An investigation of the ecology of rhizobia that nodulate white and subterranean clovers in response to soil pH

A thesis

submitted in partial fulfilment

of the requirements for the Degree of

Doctor of Philosophy

at

Lincoln University

by

Anish Sharadkumar Shah

Lincoln University

2019

Abstract of a thesis submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

An investigation of the ecology of rhizobia that nodulate white and subterranean clovers in response to soil pH

Anish Shah

White (WC) and subterranean (SC) clovers are widely used and economically important in New Zealand pasture systems. They are used as the base legume in grass-based pastures where they form an effective symbiosis with nitrogen-fixing rhizobia. Biological nitrogen fixation (BNF) by rhizobia is an important process for New Zealand agriculture, particularly in hill and high country areas where the use of nitrogen fertiliser is limited. In these environments, pH is often not optimal for plant and microbial growth. In particular, low (acidic) pH soils have a negative effect on the plant-rhizobia symbiosis and thereby reduce clover growth. Prior research has shown that pH affects the community structure, diversity and function of bacteria in the soil. However, there is limited research on the specific relationship between soil pH and the ecology of rhizobia, either when they are free-living or inhabiting the legume nodule. The overall goal of this research was to increase fundamental knowledge on the relationship between soil pH and the rhizobia nodulating subterranean and white clovers. This will improve our understanding of rhizobial ecology within the complex soil environment and contribute towards identifying strains tolerant of soil pH.

In Chapter 2 the relationship between soil pH and the diversity of bacteria that inhabit the nodules of SC and WC across soils that encompassed a broad range of pH was investigated by amplicon sequencing (*16S rRNA* and *nodC* genes). The bacterial community was assessed in 5,299 nodules recovered from SC and WC planted in 44 soils that varied in their edaphic properties, including pH. As expected, fewer nodules were formed on both clovers at low soil pH. Rhizobia comprised ~92% of the total reads in both clovers, however several non-rhizobia genera were also present. Soil pH influenced the community structure of bacteria within the nodule. The alpha diversity of nodule microbiome in SC nodules was higher than in WC nodules and SC nodules also harboured a higher relative abundance of non-*Rhizobium* bacteria than WC. Beta diversity of *Rhizobium* and non-*Rhizobium* bacteria was influenced by clover species rather than edaphic factors. Sequence analysis of the *nodC* gene was used to investigate subspecies diversity in the rhizobia community within the nodules. This work revealed 353 unique nodC protein sequences within the 5,299 nodules, and these varied in their distribution in hosts and soil pH. There was no relationship between soil pH or other soil properties on the diversity and abundance of nodC protein in the nodules of SC or WC.

In Chapter 3, pH adaptation by rhizobia strains in relation to the pH of their soil of origin was investigated. A high throughput pH bioassay was developed and optimised to test pH adaptation in a large number of strains (n = 299) in a time- and cost-efficient manner. Adaptation/tolerance by

rhizobia strains to extremes of soil pH was a rare trait (≤ 6%). The optimal growth of strains at a particular media pH correlated (r = 0.6) with the pH of the soil the strains were isolated from. However, this association was strongly influenced by the growth of strains from soils of pH >7.0, especially in strains isolated from SC nodules. Naturalised strains that are pH adapted may compete well and/or persist in soils with pH <5.0 or >7.0 compared with introduced, commercial strains. Hence, this result identified the opportunity to improve commercial inoculants through the identification of strains of rhizobia that nodulate SC and WC, fix N and also have tolerance to soil pH extremes.

Rhizobia occupy different habitats (bulk soil, rhizosphere soil and root nodules), which differ in their physicochemical conditions, and in the type of carbon sources present. Soil pH is also known to influence soil C cycles through an effect on microbial communities. For this reason, the divergent strains of WC identified from the pH assay were screened on 190 carbon sources (Biolog™) to determine their carbon utilisation profiles. This was done to determine whether pH-adaptation in some strains was linked to a competitive advantage in soils, based on metabolism of a diverse range of carbon sources. A very high degree of phenotypic variation was observed, with the 19 strains placed into 10 groups. Four of the strains used the most carbon sources and exhibited the highest growth (larger AUCs). Some carbon sources (e.g. D-glucosamine, laminarin, L-pyroglutamic acid) stimulated strain growth more than others (e.g. citric acid, tyramine, oxalic acid). A strong association was found between the phenotypic variation and genetic distance of the strains. However, there was no evidence of association of phenotypes with soil pH, soil order, sampling location, other soil physicochemical or climatic properties. Knowledge of C preference by pH adapted strains could support the production of strain-specific formulations that can stimulate maximum growth and assist the effective establishment of strains in the field.

The legume nodule is a plant structure that is critically important in the global N cycle. This research investigated the community ecology of rhizobia and "other" bacteria recovered from legume nodules in a background of variable soil pH. The results produced new information on the relationship between soil pH, the nodule biome and pH adaptation by rhizobia. The nodule microbiome was influenced by external soil pH, especially in SC. Although pH adaptation was rare, soil pH did correlate with pH adaptation by strains of rhizobia, raising the potential for selection of strains better suited to sites with sub-optimal soil pH, such as the New Zealand high country. Adaptation to pH was not explained by the ability of strains to utilise specific C-sources, however, this work did highlight the potential for improving formulations by addition of specific C compounds. The strain collection produced by this work may also contain candidate strains with potential as commercial inoculants for soil sites with pH outside the optimum.

Keywords: *Rhizobium*, *Trifolium repens* L., *T. subterraneum* L., soil pH, bacterial community, next generation sequencing, naturalised strains, pH-adaptation, carbon utilisation, Omnilog, new clover inoculants.

Dedication

I wish to dedicate this thesis to two fantastic human beings: my Granddad (Dadaji), Kantilal Bhimji and my father (Pops), Sharadkumar Shah. You are my superstars. The sacrifices you made so that I could attend University in New Zealand and build a life for myself here, so far away from home, are invaluable. I would not be who I am today if it weren't for you two.

Dadaji, you taught me strength and resilience beyond measure. Your words will resonate with me forever. I miss you.

Pops, your humility is something I will always aspire to achieve. Thank you for always supporting me and looking out for me.

Acknowledgements

I do not believe in luck, fate or destiny. What I do believe in is hard work, perseverance, empirical evidence and most of all support from friends, family and colleagues. I'm not an island, and my successes are a team effort.

The combined experience and expertise of my supervisors has been invaluable. Thank you Hayley for all your support, guidance, patience and kindness. Derrick kept me grounded and honest throughout the PhD. Steve always showed extra enthusiasm in my results and delivered feedback in record-shattering time. Thank you all for giving me this opportunity to learn from you.

My sincere appreciation goes to all the wonderful people who have given me excellent technical support, without which I'd still be three years away from completion! Drs Céline Blond and Shanika Tennakoon were paramount to the successful execution of most of the experiments. Massive thank you to: Candice Hume, Leanne Hassall, Sandy Hammond, Norma Merrick, Dalin Brown, Sandra Young, Brent Richards and Leona Meachen.
Thank you to Alasdair Noble for being my 'R mentor' and for all the hours of statistical help you've given me. Aurélie Laugraud for all your patience and bioinformatics help especially with my NGS data. Paul Maclean for stepping in when Aurélie was away. Sam Brown for the R motivation I direly needed. Thanks to Laura Villamizar, Tom Maxwell and Valerio for imparting their gems of wisdom.

The Landcorp staff were very obliging when I conducted the soil surveys. Many thanks to Scott Harpham, Jess Dunbar and all the farm managers who facilitated our soil sampling trips.

The 'admin ladies' for always smiling and being so accommodating to me. My gratitude goes to Emma Gambles, Barbara Anderson, Gail Harkerss, Amal Torky and Marianne Duncan.
A big '*Asante sana*' to the Plant-pathology team and in particular, my colleagues: JR, Drs Kathryn, Wisnu, Neeraj, Krit, Aimee, Shanika and Tasha for advice, technical help, gym buddies and sometimes just listening to my (many) rants. Also thumbs up to the 'Morning tea group' for the quality banter and laughter.

Special mentions go to my amazing friends without whom this PhD journey would have been impossible to endure. Natalia, Mike Cripps, Laura, Céline, Damian, Michael Welsh, Rachel, Tom, Becci, Eric, Lucy, Federico and Annabel. Thank you for 'putting up' with me! Also, kudos to practically everyone who has warmly indulged me throughout my time in New Zealand. You're all AWESOME!

Finally, to my family – there are not enough superlatives for me to express my gratitude. You've all shown me understanding, love and eternal support. Thank you for believing in me. I hope to continue making you proud.

Contents

Abbreviations

Al	Aluminium
AMN	Anaerobically Mineralisable Nitrogen
BNF	Biological Nitrogen Fixation
C	Carbon
Ca	Calcium
CEC	Cation-exchange Capacity
CFU	Colony Forming Units
K	Potassium
Mg	Magnesium
MPN	Most Probable Number
N	Nitrogen
Na	Sodium
NGS	Next Generation Sequencing
OD	Optical Density
Olsen P	Phosphorus
OM	Organic Matter
OTU	Operational Taxonomic Unit
PAN	Potentially Available Nitrogen
PCO	Principal Coordinates Analysis
Rlt	*Rhizobium leguminosarum* bv. *trifolii*
SC	Subterranean clover (*Trifolium subterranean*)
SDW	Shoot Dry Weight
S	Sulphate sulphur
TAE	Tris-acetate-EDTA
WC	White clover (*Trifolium repens*)
YMA	Yeast Mannitol Agar
YMB	Yeast Mannitol Broth

List of tables

List of figures

1 Literature review

1.1 Pastoral farming

New Zealand's economy is heavily reliant on the pastoral sector, and legumes make up a sizeable proportion of all pastures. A report from 2012, commissioned for the Pasture Renewal Charitable Trust (PRCT), estimates the contribution of pasture-based products to the New Zealand economy at approximately 12% ($24.5 billion, NZD) of total GDP (Business and Economic Research Limited, BERL 2012). Different types of pastoral farming have become specialised in different regions across New Zealand. Much of the easier-contoured and flat land in New Zealand has been irrigated and converted to dairy farming, particularly in Waikato, Canterbury, Otago and Southland. Consequently, dryland sheep and beef farming has intensified in the hill and high-country regions of both main islands (Moot 2013). According to Statistics New Zealand (2012), around 2.3 million hectares (M ha) were devoted to dairy cattle and 4.3 M ha to sheep and beef cattle grazing. Pastoral farming thus occupies more than 46% of the total agricultural land in New Zealand (www.stats.govt.nz/yearbook2012). The predominant form of pasture in the dairy areas is a mixed system composed of perennial ryegrass (*Lolium perenne* L.) and white clover (*Trifolium repens* L.). In hill country a more diverse range of pasture species are found which includes subterranean clover (*T. subterraneum* L.) in summer dry regions.

1.2 Pasture legumes in New Zealand

Many of the world's major food and feed crop species are legumes, such as soybean, pea, clover, chickpea and alfalfa (lucerne). Globally, legumes are the second largest group of food and feed crops. Legumes are responsible for more than 25% of the world's primary crop production with 247 million tons of grain legumes (those grown for their seed) produced annually (Ferguson et al. 2010). The New Zealand pastoral sector is more reliant on legume-based production than any other country (Charlton 2012). A number of legume species are grown as pasture depending on their adaptation to soil types, most of these being different clover species such as red (*T pratense* L.), Caucasian (*T. ambiguum* M. Bieb.), subterranean, balansa (*T. michelianum*) and white clovers. In addition, lucerne (*Medicago sativa* L.) and lotus (*Lotus pedunculatus*) are also grown in some areas, particularly as dryland pastures (Moot 2013).

The genus *Trifolium* is one of the largest in the *Fabaceae* family and is grown widely in improved pasture systems throughout cool temperate regions (Melino et al. 2012).

1.2.1 White clover

Of the legumes, white clover (WC) is most widely used in New Zealand, and has a financial contribution of fixed nitrogen to the pastoral sector at a value estimated to be in excess of $3 billion per annum (Caradus et al. 1996). WC has been used and bred in New Zealand since the

1930's, and after an intensive breeding program to improve varieties, the cultivar 'Grasslands Huia' was developed (Caradus 1989). This WC cultivar has dominated world markets since 1964 due to its adaptability, performance and competitive price (Caradus 1989; Woodfield and Easton 2004). WC is often sown together with perennial ryegrass, and grows well on fertile soils; but, like many plants, it struggles to establish in dry areas. Despite its importance and wide usage, WC is a difficult species to work with because it demonstrates genotypic plasticity, i.e. there is a high level of genetic variation both within and between WC populations (Voisey et al. 1994). White clover is allogamous (outbreeding) and heterozygous, so a cultivar is a tamed population enriched for a subset of alleles that confer the traits in that cultivar (i.e. leaf size, growth habit, persistence), whereas the rest of the genome is quite diverse (Andrew Griffiths, 2018 pers comms).

WC requires an adequate supply of nutrients such as phosphorus and sulphur to maximise its ability to fix nitrogen (N) (Caradus 1989). To fix N, it forms a symbiosis with strains of the bacterium *Rhizobium leguminosarum* bv. *trifolii* (*Rlt*). WC growth and N-fixation rates are also dependent on seasonal variation in soil temperature and moisture, the strains of *Rlt* which colonise it and management factors which affect its competitive ability. In permanent pastures, with 20% of the total dry matter composed of WC, N-fixation for total pastures is likely to average 100 kg N ha^{-1} $year^{-1}$ (Ledgard 2001). In the absence of mineral nitrogen, there is a direct relationship between N-fixation and WC growth, but increasing the supply of exogenous mineral nitrogen reduces N-fixation (Caradus et al. 1996).

1.2.2 Subterranean clover

Subterranean clover (SC) is also sown in mixed stands, generally with perennial ryegrass or cocksfoot (*Dactylis glomerata* L.) but can also be oversown or direct drilled as the sole legume into dry hill country areas of New Zealand. It is an annual legume and advocated for east coast, summer dry regions particularly on north and west facing slopes of the foot hills in the North and South Islands (Charlton 2012; Teixeira et al. 2015). It reproduces predominantly by self-fertilisation, and thus, cultivars of SC retain their true-breeding characteristics even when grown in mixtures (Suckling et al. 1983; Smetham 2003). As a consequence a SC cultivar is more uniform than a WC cultivar.

The geographic distribution of the three sub species of SC is largely determined by edaphic factors. *T. subterraneum* cv. *subterraneum* and *T. subterraneum* cv. *yanninicum* prefer acid to neutral soils, while *T. subterraneum* cv. *brachycalycinum* can tolerate higher pH and alkaline soils (Smetham 2003). Similar to WC, nodulation of SC is carried out by strains of *Rlt*. However, there are distinct differences between strains which are effective and ineffective at nodulating these two different clovers. Some strains effectively nodulate either SC or WC, which is related to strain compatibility in association with the host (Robinson 1969).

1.3 Rhizobia

Rhizobia is the generalised name of the group of Gram-negative, α-*Proteobacteria* known to nodulate legumes. Rhizobia are facultative symbionts, and they survive in a complex environment alongside a diverse microbial community by adopting an oligotrophic lifestyle (Poole et al. 2018). In the absence of a suitable host, rhizobia can exist as free-living soil bacteria where they compete and exist in the soil alongside the diversity of other microflora. When in symbiosis with legume roots, they are able to fix atmospheric N (biological nitrogen fixation, BNF) directly via infection pockets, called nodules (Figure 1.1). The legume host provides the bacterium with carbon, energy and a controlled environment in which to live.

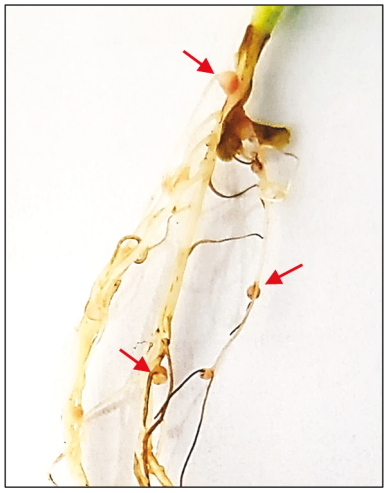

Figure 1.1: White clover root (~42 d old) showing pink nodules (red arrows).

The evolution of a large bacterial genome (~5–10 Mb) was likely enabled by such a dramatic life history change, and allows rhizobia to encode several genes which are important for endosymbiosis (Poole et al. 2018). Rhizobia are categorised according to two characteristics: i) their ability to produce acidic or alkaline compounds when grown on suitable media, and ii) the location of the nodulation (*nod*) and nitrogen-fixing (*nif* and *fix*) genes (Sprent 2009). Rhizobia are found in seven bacterial families and divided into 15 genera. Within the Rhizobiaceae family there are 43 species described in the genus *Rhizobium*, 13 in *Sinorhizobium*, 3 in *Neorhizobium* and 1 in *Shinella* (Giller et al. 2016). Some species of rhizobia have a taxonomic sub-division called biovar (sometimes referred to as symbiovar) and there are currently numerous symbiovars in the genera *Sinorhizobium*, *Rhizobium*, *Mesorhizobium* and *Bradyrhizobium* (O'Hara et al. 2016).

1.4 Nitrogen fixation

Legumes, in symbiosis with rhizobia, have the ability to fix atmospheric nitrogen into a biologically available form. The interaction between rhizobia and legumes is generally host-specific, wherein the range of hosts which a particular species/strain of *Rhizobium* can nodulate can be narrow or broad (Young and Johnston 1989; Dénarié et al. 1992; Amarger 2001). A successful legume-

rhizobia symbiosis results in BNF in the legume roots. Biologically fixed nitrogen improves sward quality, has seasonal complementarity with grasses and improves feed intake and utilisation rates (Caradus et al. 1996). Approximately 2.5 x 10^{11} kg of ammonia is fixed from the atmosphere through BNF (legumes and cyanobacteria), globally per annum (Rao 2014). Currently, about twice as much industrially fixed nitrogen is needed as fertiliser for crop production to equal the effects of BNF, mainly because of losses due to leaching. Furthermore, the demand for N fertiliser is projected to increase by 1.5% globally, by 2018 and a potential deficit in N fertiliser has been forecast for the Oceania region (FAO 2015 – World Fertiliser Trends and Outlook to 2018). Legumes can fix up to 80% of their own N and, in addition, can contribute to the yield of subsequent or companion crops through the release of fixed N into the soil (Rao 2014). This offsets some of the need to apply nitrogen fertiliser, and has allowed New Zealand to develop relatively low-cost, year-round, pasture-based farming systems.

The integral place of WC and SC in New Zealand pastures ensures that New Zealand farmers are highly reliant on BNF. Both species contribute inputs of around 35 Kg N $ha^{-1}yr^{-1}$ per tonne of legume grown (Parfitt et al. 2006). This high dependence of BNF has caused a reduction in the requirements for input of mineral N fertiliser to maintain pasture productivity. Despite this, New Zealand's N fertiliser use is changing rapidly, increasing from 5 x 10^4 tonnes in 1989 to 3.4 x 10^5 tonnes in 2003, thus increasing the load of N in the New Zealand environment (Parfitt et al. 2006).

1.4.1 Legume-rhizobia symbiosis

The presence of rhizobia is not always sufficient to ensure optimal nitrogen-fixing capacity in the host legume, since rhizobial strains differ in their efficacy (often termed symbiotic potential), and the symbiosis is further regulated by biotic and environmental conditions (Slattery et al. 2001; Zahran 2001; Ballard et al. 2004). There are differences in the amount of N-fixation activity among rhizobia and also in different rhizobia-legume symbioses. The rates of BNF by naturalised rhizobia in New Zealand differs between 2 and 100% compared with a highly effective control (Rys and Bonish 1981), and nearly all rhizobia-legume symbioses do not realise the full potential of an efficient partnership (Denton et al. 2002). The symbiotic relationship between rhizobia and legumes is relatively well understood. A molecular dialogue initiates the colonisation of the root hairs by the bacteria and initiation of symbiotic nitrogen fixation in the roots. The infection process is initiated in the epidermal root hairs of the legume, where the secretion of isoflavonoids by the host plants induces the expression of *nod* genes (e.g. *nodABC*) in the bacteria. *Nod* genes are involved in the synthesis of lipochito-oligosaccharide Nod factors (e.g. NodD) which trigger early symbiotic responses such as the curling of plant root hairs and division of meristematic cells. The rhizobium becomes entrapped within the 'Shepherd's crook' (Figure 1.2), forming a small rhizobial colony and nodule organogenesis is initiated. This in turn induces cellular changes causing the formation of infection threads which release the bacteria into the plant cells and

differentiate into symbiotic forms called bacteroids (Fischer 1994; Cooper 2007; Oldroyd et al. 2011; Wielbo et al. 2012; Haag et al. 2013; Janczarek et al. 2015).

The bacteroid nitrogenase enzyme complex reduces atmospheric nitrogen to ammonium inside the nodule's anaerobic environment, which protects the nitrogenase from inactivation by oxygen (Janczarek et al. 2015). The type of nodules formed on *Medicago* and *Trifolium* species are referred to as indeterminate. These have a persistent meristem which continually provides new nodule cells for subsequent bacterial infection and these nodules are characterised by their cylindrical shape and multiple developmental zones (Ferguson et al. 2010). DNA replication of the entire bacterial genome in indeterminate nodules occurs in the absence of cell division, leading to much larger genomes than those in free-living bacteria (Barnett and Fisher 2006).

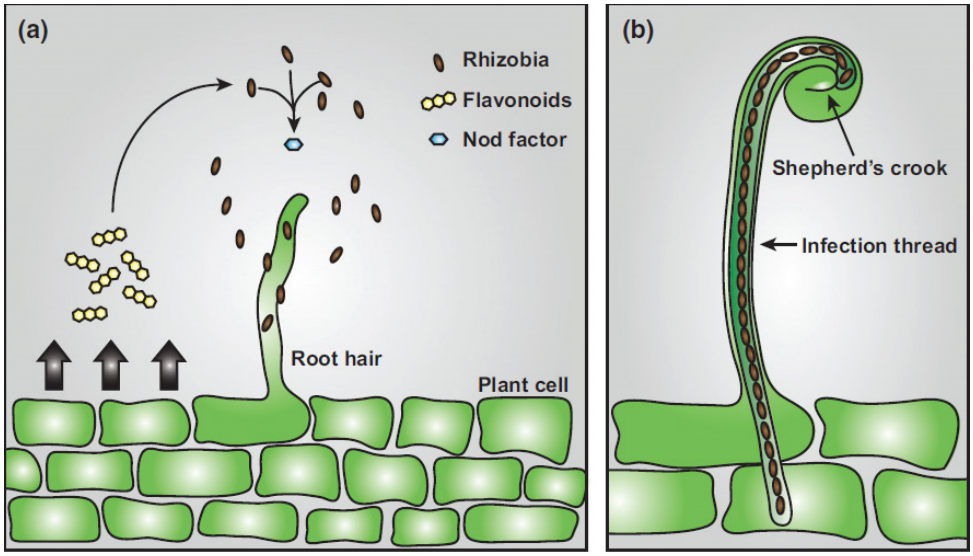

Figure 1.2: Rhizobia interacting with legumes. (a) The legume secretes flavonoids which induce the rhizobia to produce Nod factors and attract them to the plant root hair cells. (b) Nod-factor signalling triggers root hair curling which traps the rhizobia in Shepherd's crooks. Inward growth of the root hair tip forms infection threads, which allow the rhizobia to enter the cortical cell layers of the plant root (Haag et al. 2013).

Symbiotic nitrogen fixation genes have been broadly divided into *nod*, *nif*, and *fix* genes (Fischer 1994). The *nod* and *nif* gene regions have been cloned and mapped in a *Rlt* symbiont (Scott and Ronson 1984) and *fixABC* genes have been analysed in *S. meliloti* (Earl et al. 1987). To be transcribed, *nod* genes require inducers released by the host plant and the transcription activator, NodD (Long 2001). In the case of *R. leguminosarum* and *S. meliloti*, the *nod*, *nif* and *fix* genes are located on the mega-plasmid, called a symbiotic plasmid (pSym) (Janczarek et al. 2015).

1.5 Factors affecting the legume-rhizobia symbiosis

The community of rhizobia in the soil and the way in which they interact with the plant is affected by different biotic and abiotic factors (Garbeva et al. 2004; Howieson and Ballard 2004). Thus, it is

important to gain knowledge on how biotic, abiotic and edaphic (soil-specific) factors affect rhizobia populations in the plant (nodule) and soil. This may also assist in identifying commercially useful strains of rhizobia which are tolerant to abiotic and edaphic stresses and more able to compete and persist in the free-living phase. Furthermore, tolerance of key abiotic, biotic and edaphic factors may confer an advantage to the rhizobia during the establishment phase, where the bacteria are delivered to soil (either as coated seed, granule, or other formulation) and interact with a compatible host legume root system. Some examples of the biotic factors that influence rhizobial communities, and in turn the legume-rhizobia symbiosis include the general soil microflora (intra- and interspecific competition) and the species of host legume. A range of edaphic factors affecting rhizobial survival and nodulation efficiency have been identified. These include: acidity (aluminium toxicities), alkalinity (calcium and boron concentrations), aridity, salinity, heavy metals, soil structure and soil mineral (nitrogen, phosphorus, carbon, sulphur) concentration (Slattery et al. 2001; Howieson and Ballard 2004).

1.5.1 Soil pH

In general, bacterial richness and diversity in the soil varies across ecosystem types and in particular is affected by soil pH, with acid soils producing very low overall diversity and richness (Fierer and Jackson 2006). The productivity of leguminous pastures is significantly affected by natural soil acidity and by accelerated acidification due to increasing industrial pollution and agricultural practices (Dilworth et al. 2001). The failure of N-fixing *Rhizobium*-legume symbioses in acid soils is a significant problem affecting agricultural production in many areas of the world, and is due to poor growth and survival of rhizobia and clovers under these conditions (Watkin et al. 2000). Soil pH influences rhizobia strain variation (Watkin et al. 2000; Slattery et al. 2001; Howieson and Ballard 2004). Watkin *et al.* (2000) conducted a field study with *T. subterraneum* to determine the performance of six strains of *Rlt* in an acid soil (pH_{CaCl2} 4.2). Their results showed that the strains exhibited strong variation in their ability to nodulate plants, with some strains significantly better at nodulating plants in the acidic soil. The six strains differed in their ability to persist in the acid soil over a two year period in the absence of a host plant (clover). They identified strains WU95 and WSM409 as ones with potential for improving the production of clover on acid soils.

In a study by Brockwell *et al.* (1991), pH affected the abundance of *S. meliloti* (Table 1.1) where a low abundance was demonstrated in acidic soils, even when the frequency of the host plant was high (Brockwell et al. 1991). *Medicago* spp. occur frequently on alkaline soils, but the number of plants decreases as the soil pH falls (Brockwell et al. 1991).

Table 1.1: Mean values for soil pH, number of *S. meliloti* and medic frequency for seven soil groups of the Macquarie region, New South Wales (adapted from Brockwell *et al.*, 1991).

Soil group	No. of sites	Average soil $pH_{(water)}$	No. of *S. meliloti* [log_{10}(MPN +1)/g]	Medic frequency (%)
Black earths	14	7.12	4.15	75
Grey and brown soils	18	6.60	3.92	94
Red loams	11	6.37	2.80	58
Red-brown earths	22	6.12	2.47	87
Brown acid soils	9	5.86	2.99	88
Red and yellow solodic soils	6	5.75	0	6
Sandy solodic soils	4	5.67	1.88	22

Other studies have also shown that the *S. meliloti–Medicago* spp. symbiosis is sensitive to soil acidity, and the growth of most *S. meliloti* strains, as well as the process of nodule initiation, is inhibited below soil pH_{CaCl_2} 6.0 (Munns 1968; Rice et al. 1977). The impairment of the nodulation process, specifically the steps preceding root hair curling, is the first symptom observed with declining pH and most likely the primary limitation to lucerne growth below pH_{CaCl_2} 5.0 (Munns 1968). Charman *et al.* (2008) performed experiments with lucerne in solution cultures maintained at pH 5.0 and identified strains of rhizobia which were able to successfully nodulate the plants. In particular, strain SRDI291 was found to nodulate a greater percentage of seedlings (85%) as well as producing more nodules (3.5) per plant compared with the commercial strain (Figure 1.3) (Charman et al. 2008). However, when tested in soil, the SRDI291 strain did not nodulate proportionately more lucerne plants than strain RR1128 (Humphries et al. 2009). This later work found that compared with strains SRDI291 and RRI128, strain SRDI675 was more effective at nodulation both in solution culture as well as in soil (pH_{CaCl_2} 5.5).

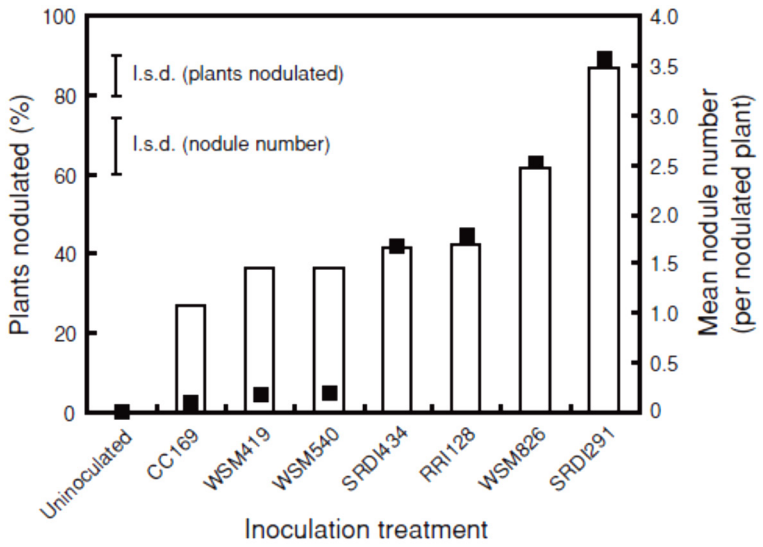

Figure 1.3: Effect of inoculation treatment on the percentage of lucerne plants nodulated (black blocks) and the number of nodules per plant (bars). RRI128 is the commercial strain (Charman et al. 2008).

Attachment of rhizobia to legume roots is also affected by soil pH. At acidic pH, *R. leguminosarum* uses glucomannan at its cell pole to attach to lectins on the root hairs and surface of pea. However, glucomannan is not involved at alkaline pH, probably because plant lectins are lost at high pH (Williams et al. 2008; Poole et al. 2018).

A study carried out on faba bean (*Vicia faba*) revealed that the mass, number of nodules formed and the activity of nitrogenase was significantly reduced as soil alkalinity increased. The researchers found that an increase in pH_{H2O} from 7.5 to 9.0 reduced nodulation to 50%, but this detrimental effect was reduced when plants were inoculated with both arbuscular mycorrhizal fungi and rhizobia (Abd-Alla et al. 2014). Thus, it is likely that P-deficiency in the alkaline soil was a significant constraint for the host legume to support the symbiosis.

1.5.2 Other abiotic factors affecting rhizobia

Soil Nitrate – The concentration of nitrate in soils can also affect which strains of rhizobia are effective at colonising legume roots and fixing nitrogen (Herridge et al. 1984; Unkovich and Pate 1998; Howieson and Ballard 2004). Some symbioses can delay nodulation in the presence of soil nitrate and then quite adequately proceed to nodulate when the soil reserves of N are diminished (Howieson and Ballard 2004). This is particularly important for dairy pastures in New Zealand, where N is usually added early during establishment of the ryegrass/white clover pasture. If this N addition is not followed quickly by substantial rains that leach N from within the rooting zone of germinating legumes, nitrate can cause significant disruption to nodulation (Howieson and Ballard 2004). There are some reports for legumes grown under managed or controlled conditions that the proportion of total plant N obtained from N-fixation decreases with increased soil N availability (Novák et al. 2009; Liu et al. 2010; Andrews et al. 2011).

Soil Carbon – Rhizobia have the ability to survive both in the presence and absence of legume hosts, as such, they can metabolise different carbon sources both inside the nodule and in the rhizosphere soil. Variation in the utilisation of various carbon sources may provide some free-living rhizobia a competitive advantage in the harsh/dynamic soil environment. Carbon availability in soils is also strongly influenced by pH (Andersson et al. 2000; Rousk et al. 2009). Plasmid-borne genes confer significant metabolic versatility to rhizobia, which is important for their nodulation competitiveness and adaptation to various environments (Zahran 2017). Several of those plasmid-borne genes help rhizobia to utilise various substrates such as adonitol (sugar alcohol present in bulk soil), arabinose (present in root exudates), catechol (aromatic compound in bulk soil), inositol (vitamin in either root exudates or soil), lactose (bulk soil), and malate (root exudates) (papers cited in Zahran, 2017). A number of studies (reviewed in Stowers, 1985) showed differences in carbon metabolism between fast- and slow-growing rhizobia species and/or strains. Some evidence from those studies suggests that fast-growing rhizobia can utilise a

variety of carbon substrates, whereas the slow-growers can only utilise relatively fewer carbon sources (Stowers 1985).

Temperature – This is a major abiotic factor that influences all biological processes, due to enzyme sensitivity to extremes of temperature. Numerous studies (reviewed in Alexandre and Oliveira (2013)) have examined the effects of temperature stress on rhizobial growth, survival, nodulation and nitrogen fixation in legumes. High temperatures have an inhibitory effect on adherence of bacteria to root hairs, as well as root hair and infection thread formation (Alexandre and Oliveira 2013). Low temperatures are equally detrimental to nodulation and N-fixation, delaying and even inhibiting these processes in various legume-rhizobia symbioses (Zhang et al. 1995). However, most of these temperature-stress studies have been done in controlled laboratory conditions, and it is important to consider that growth conditions in the laboratory are usually very distinct from the natural conditions experienced by bacteria in the soil. Hence, temperature tolerance of rhizobia may also depend on other environmental conditions not easily reproduced in laboratory studies (Alexandre and Oliveira 2013).

Salinity – This interferes with various phases during initiation of the *Rhizobium*-legume symbiosis, nodule formation and nitrogen fixation. Four steps have been identified where salinity can interfere with root hair nodulation. These steps are: root hair formation, signal exchange between the plant and rhizobia, the physical attachment and infection thread formation, and lastly the formation and functioning of the nodule (Bruning and Rozema 2013).

Nutrient deficiency – This can affect both free-living rhizobia as well as symbionts engaged in nodulation (reviewed in O'Hara (2001)). Changes in soil pH affects the solubility and availability of some inorganic nutrients, and may also induce nutrient deficiency in some cases (O'Hara 2001). Divito and Sadras (2014) confirmed that deficiency of phosphorus, potassium and sulphur reduce nodule growth, number and activity more than both shoot and nodule mass, which indicated a reduction in nodule productivity. O'Hara (2001), mentioned that K, Ca and Mg form part of the essential macro elements for rhizobia. Furthermore, the presence of organic matter may also influence nutrient availability (O'Hara 2001). This thesis will focus on P, K, Ca, Mg, Al, and organic matter because of the relationship between soil pH and those physicochemical properties and due to their effects on rhizobia growth and survival as outlined by O'Hara (2001).

1.5.3 Biotic factors affecting rhizobia

Competition for nodule initiation is most commonly intraspecific, i.e. between strains of a single species of root-nodule bacteria (Thies et al. 1991). However, interspecific competition plays a role when a particular legume (e.g. *Phaseolus vulgaris*) is nodulated by more than one species of rhizobia (Howieson and Ballard 2004). The *Rhizobium*-legume interaction is improved when arbuscular mycorrhizal fungi (AMF) are associated with the symbiosis, due to the acquisition of N

and P from soil. In general, the additive effects on photosynthesis are improved when the plant is in symbiosis with both rhizobia and AMF (Kaschuk et al. 2009).

In addition to inter- or intraspecific competition, the distribution of rhizobia in soil may also influence nodulation efficiency. For example, the soil rhizobial population may occupy the soil profile where the roots penetrate, whereas the inoculant strains may remain concentrated around the seeds (López-García et al. 2002). López-García *et al.* (2002) reported that the position of *Bradyrhizobium japonicum* strains in the soil was a strong determinant for nodule occupancy in soybean. Sanctions by legume hosts also have potential to influence the symbiosis. Kiers *et al.* (2006) (and studies cited therein) suggested that legumes hosts are able to differentially affect the fitness of rhizobia in nodules based on their N-fixation rates, perhaps by varying resource allocation. Ramana (2018) found that older (low selective breeding) white clover cultivars employ sanctioning to select for more effective rhizobia symbionts.

The nodule microbiome is occupied by both rhizobia and "other" bacteria (Hartman et al. 2017; Wigley et al. 2017), irrespective of whether the nodule dwellers are involved in nodule formation or nitrogen fixation. However, the functional significance and roles of these other taxa inside nodules, in response to the external environment is unclear. It has been shown in several studies (reviewed in Martinez-Hidalgo and Hirsch (2017)) that nitrogen-fixing rhizobia live alongside other bacteria in legume nodules. *T. pratense* (red clover) was co-inoculated with rhizobia and a mixture of *Pseudomonas*, *Janthinobacterium* and *Microbacterium*, and resulted in increased shoot weights (Hartman et al. 2017).

1.6 Rhizobia inoculants in New Zealand pastures

In New Zealand, the practice of inoculating legume seed with rhizobia first started in 1927 using agar cultures of *Rhizobium* (Hastings and Drake 1960). Since 1955, commercially prepared inoculants have been available and coated or inoculated seed has been marketed since the early 1960's. Australian inoculants entered the New Zealand market in 1974 (MacKinnon et al. 1977), and since then most of our important commercial rhizobial strains are of Australian origin.
The commercial strains of rhizobia inoculants used in Australia and New Zealand are TA1 for white clover and RRI128 for lucerne (Unkovich 2012). The commercial strains of rhizobia used for SC pastures were previously WU95 (Watkin et al. 2000) and WSM409 (Gemell et al. 2005), but the current strain in use is WSM1325 (Yates et al. 2008; Nangul et al. 2013). Strain TA1 is the most widely used commercial inoculant for WC. However, a number of field isolates have been recovered which are more effective at N-fixation than TA1, especially under stressful soil conditions (Riffkin et al. 1999; Watkin et al. 2000; Unkovich 2012). Thus, it is likely that new commercial inoculants could be developed as more research is carried out on the efficacy of rhizobial strains under different soil conditions and for different legume species.

To realise the potential benefits of BNF through commercial inoculation in harsher soil environments, it is crucial to find those rhizobial strains which can tolerate abiotic stresses, and confer significant benefits in terms of nodule occupancy (infectiveness) and the amount of nitrogen fixed (effectiveness) in the legumes under harsh conditions. There are four possible interactions between *Trifolium* spp. and their *Rlt* symbionts: (i) no symbiotic interaction (zero infectiveness), (ii) ineffective or parasitic interaction, (iii) a partially effective interaction (some N-fixation), and (iv) an effective symbiotic interaction (high N-fixation) (Yates et al. 2003; Howieson et al. 2005). The use of laboratory growth media can provide an opportunity to study the effects of specific soil toxicities or deficiencies on rhizobium growth and survival. However, care must be used when selecting rhizobium strains for acid tolerance based on laboratory data and extrapolating this to field conditions (Slattery et al. 2001).

Soil pH is a major determinant to establish effective rhizobia-legume symbioses. Therefore, experimental approaches should focus on the identification of strains of rhizobia with greater tolerance to pH extremes, and which also nodulate and fix nitrogen at high rates. This should ensure their successful establishment and growth of pasture legumes in diverse soil habitats. In New Zealand, commercial strains are required to i) colonise legumes with high numbers of effective rhizobia and to out-compete ineffective natural rhizobia populations; and ii) to bolster populations which have limited persistence due to adverse soil conditions (Deaker et al. 2004). There is also a discrepancy between the rapid development of new legume cultivars and identification of effective commercial rhizobial strains to inoculate those legumes. This is evidenced by the current use of commercial strains of rhizobia identified more than five decades ago (e.g. strain TA1). There exists a significant opportunity to develop new inoculants for clover, especially by employing cutting-edge technologies to identify efficacious strains.

1.7 Using high-throughput technologies

1.7.1 Next Generation Sequencing

Next Generation Sequencing (NGS) approaches have the potential to provide a greater depth of understanding about the nodule community structure in different soil backgrounds. NGS generates huge amounts of data, factoring in the different treatments, plants and also individual nodules. Application of this high-throughput technology has allowed greater depth in bacterial community analysis (Shokralla et al. 2012; Taberlet et al. 2012; Logares et al. 2014). The *16S rRNA* gene is the most common target to assess the taxonomic diversity of complex bacterial populations. Recent studies using *16S rRNA* gene sequencing from various ecosystems showed that soil pH was the strongest predictor for community structure, diversity and function (Lauber et al. 2009; Kaiser et al. 2016). From NGS data, the alpha diversity (richness of genotypes) can be calculated to determine if overall diversity is greatest at neutral pH and declines at low and/or high pH (the intermediate disturbance hypothesis - (Connell 1978)). Using beta-diversity

measures (Anderson et al. 2011) from the same data set, community composition can be compared across samples, and the role of pH and other soil factors defined as a driver of community composition.

1.7.2 Phenotype Microarray

Phenotype microarray (PM) technology allows culturing of isolates in small volumes and provides the opportunity to perform many parallel assays in a compact space. It works by culturing cells and assaying growth colormetrically on the Omnilog unit (Biolog Inc., Hayward, CA). The Biolog microplates can be used to perform assays on utilisation of carbon, nitrogen, phosphorus, sulphur and amino acid sources by the selected strains and tolerance to different pH, nitrate and NaCl conditions (Mazur et al. 2013). Phenotypic changes in the strains can be detected by comparing growth in these plates, as environmental conditions are changed. The plates are prepared by the user, giving control and knowledge of media components (Borglin et al. 2012).

Microwell phenotype systems enable rapid screening of a wide variety of growth conditions and commercially available plates contain a large suite of substrates, nutrient sources, or inhibitors. These can provide an assessment of the phenotype of an organism, spanning C preference, autotrophy, ability to utilise macro and micronutrients, testing sensitivity to pH, antibiotics, among others. (Borglin et al. 2012).

1.8 Rationale and Objectives

The legume nodule is a specialised plant structure in which rhizobia fix atmospheric nitrogen and in doing so contributes to global N cycles. Soil pH is a first-order factor influencing the niche preferences of soil microorganisms, and there have been some examples of rhizobia adaptation to soil pH extremes (Howieson et al. 1988; Brockwell et al. 1991; Denton et al. 2002; Watkin et al. 2003). However, there is no information on whether soil pH affects the community diversity of rhizobia and other bacteria in the nodule. Thus, the overall goal of this thesis was to examine the effect of soil pH on the ecology of rhizobia and bacterial communities in the nodule.

Natural soils, with established history of soil pH, will be used as the source of bacteria for clover nodulation. NGS technology (Illumina HiSeq) will allow the bacterial communities in multiple nodules, plants, and samples to be investigated in parallel, and the relative richness and relative abundance of nodule inhabitants to be calculated. This work is novel as no research group has previously used an NGS approach to assess the relationship between nodule microbial populations and soil pH. Using two clover species that are important to NZ will provide a comparative aspect to this study.

This research has been divided into three research objectives:

1) To determine whether there is a relationship between soil pH and the nodule microbiome (Chapter 2).
   - Illumina based amplicon sequencing of the *16S rRNA* gene will be used to assess the relative abundance of microbial taxa in the nodule.
   - The genotypic diversity of rhizobia in the nodules will be assessed by amplicon sequencing of the *nodC* gene.
   - The alpha and beta diversity of nodule communities will be compared between clover species.

2) To determine whether soils contain strains of rhizobia that demonstrate adaptation to pH (Chapter 3)
   - Soil baiting with WC and SC will be used to create a collection of rhizobia from soils that have different pH.
   - A high throughput bioassay will be developed to identify pH adaptation in the rhizobia.
   - The relationship between optimum *in-vitro* pH and pH of soil of origin of stains will be assessed.

3) To determine whether pH-adapted strains have a phenotypic relationship with the ability to metabolise soil C in terms of the number or type of C-sources they can metabolise. (Chapter 4)
   - The ability of pH-adapted rhizobia to utilise 190 C-sources will be assessed using Biolog™ plates.
   - C-utilisation profiles of strains will be compared with geography, edaphic properties and the strain genotypes.
   - To determine if soil pH strongly affects microbial communities and soil C cycles.

2 Bacterial community diversity in the nodules of subterranean and white clover grown in soils of differing pH

2.1 Introduction

Since their introduction, rhizobia have spread widely throughout New Zealand and pastoral soils now contain high levels of resident rhizobia capable of nodulating subterranean, white and other clovers (Lowther and Kerr 2011). High numbers of naturalised rhizobia exist in many soils across New Zealand. Research in New Zealand and globally on clovers as well as other legumes, has shown that both environmental and edaphic factors have potential to influence diversity of rhizobia genotypes in soil and that rhizobia populations differ across spatial scales (Graham 2008; Wakelin et al. 2018). However, few studies have investigated the connection between edaphic factors and the bacteria that inhabit the nodules (Leite et al. 2017; Seth 2017).

With the recent advances in the plant microbiome it has become clear that plants can recruit specific bacterial taxa as endophytes to assist in alleviating abiotic stress, such as salinity (papers cited in Martinez-Hidalgo and Hirsch (2017)). The legume nodule is a highly specialised structure in which rhizobia undertake nitrogen fixation. Recent work has shown that the in addition to rhizobia the nodule can contain bacteria from other genera (Wigley et al. 2017). However, the understanding of whether this is a deliberate process influenced by the plant or a random process that occurs during nodule formation is unclear. For example, the *16S rRNA* diversity of bacterial communities associated with root nodules of cowpea (*Vigna unguiculata* L. Walp) grown in Brazilian soils found that the non-rhizobial communities were influenced more by the soil type than the plant genotype (Leite et al. 2017). Work done in New Zealand pasture legumes using high-throughput sequencing technologies showed that diverse bacterial taxa occupied the root nodules of lucerne (*Medicago sativa* L.) (Wigley et al. 2017), however, occupancy was not associated with edaphic factors.

The more recent application of new technologies such as metabarcoding using high throughput sequencing platforms e.g. Illumina MiSeq and HiSeq, has allowed greater depth in bacterial community analysis (Shokralla et al. 2012; Taberlet et al. 2012; Logares et al. 2014). Community analysis of bacteria can provide information about how edaphic factors affect the local diversity and abundance of functional taxa in plants and in soil (Fierer et al. 2012). These new tools offer the opportunity to assess changes in bacterial community diversity within the nodules of key pastoral legumes. The genetic diversity (*sensu* species richness) and abundance (species frequency) of complex bacterial populations in soil can be studied at a broad level using taxonomic genes such as the *16S rRNA* or targeting a subset of organisms using functional gene (e.g. *nod* genes) approaches (Sarita et al. 2005). The *16S rRNA* gene is the most common target in research assessing the phylogenetic diversity of complex bacterial populations, with recent work showing that edaphic factors strongly affected soil bacterial community structure and function,

and pH being the strongest predictor for community structure, diversity and function (Lauber et al. 2009; Kaiser et al. 2016). However, the effect of edaphic factors, such as pH, on microbial communities inside the clover nodule biome is currently unknown. Assessing a functional gene such as *nodC*, which is specific for rhizobia, may further validate the results based on *16S rRNA* gene analysis, if it is observed that the richness and/or abundance of rhizobia in nodules decreases at pH extremes. The *nodC* gene is a common nod gene essential for nodulation in many rhizobial species (Laguerre et al. 2001). The gene encodes an N-acetylglucosaminyltransferase which is involved in the first step of Nod factor assembly, and it (as well as *nodA*) is also a determinant of host range (reviewed in Perret *et al.* (2000)).

Of the pastoral legumes used commercially in New Zealand, SC and WC are good model systems in which to assess nodule bacterial community structure in response to a soil pH gradient, due to their economic importance and wide use. SC is predominantly sown in the drier hill country, whereas WC is better suited to wetter soils. The comparative aspect in nodule communities between two closely related legumes, which have different optimum growing conditions and different histories of selective breeding, will identify whether the hosts display similar or divergent patterns of belowground microbial partnerships. It is possible that clovers growing in a stressful environment (e.g. very acidic or alkaline soils) may recruit a greater diversity of bacterial taxa into its nodules to mitigate the effects of the adverse edaphic conditions.

The goals of this chapter were to: 1) determine whether there was a relationship between soil pH and the diversity of bacteria inhabiting the nodules of SC and WC; and 2) to compare between the nodule community structure of SC and WC. This work sampled soils from a range of sites across New Zealand to understand the ecology of rhizobia and other taxa within the nodule. The samples were taken from geographically diverse sites that had not been inoculated with rhizobia for ≥ 5 years and for which there was historic data of soil pH. The established nature of these soils allowed stable changes in rhizobia phenotype and bacterial communities to be investigated. A dual-target gene approach was taken in which the diversity of both the *16S rRNA* and *nodC* genes was assessed. Thus, this chapter tested the hypothesis "*that the pH of a soil affects both the total bacterial community diversity and the genotype of rhizobia within the plant nodule*".

2.2 Materials and methods

2.2.1 Soil surveys

Soils were collected from farms owned by Landcorp Farming Limited across different locations in New Zealand. From the farms, paddocks were chosen with no history of rhizobia inoculation at least five years prior to sampling, that had a permanent pasture of ryegrass-clover and with a range of soil $pH_{(water)}$[1] from the soil test data provided by Landcorp Farming Limited. Soil collection took place between October and November 2015. Fourteen farms were visited spanning the North (5 farms) and South (9 farms) Islands of New Zealand. At each farm, between two and eight paddocks were sampled, resulting in a total of 44 soil samples (Figure 2.2). The top 10 cm of the soil was sampled every 1–2 m across a 100 m transect from random locations within each of the paddocks using a 5 cm diameter corer and avoiding stock faecal matter (Figure 2.1). The soils were sieved to remove stones and root material using a 3.5 mm sieve and an approximately 200 g sub-sample sent to Hill Laboratories for chemical analysis to test different soil physicochemical properties (Chapter 2 supplementary results). Pearson's correlations of the soil properties with pH are presented in Table 2.8. After collection, sieving and sub-sampling the soils were stored in a cold room at 7°C.

Figure 2.1: Photograph of the soil core used to collect the soil and a bag with soil cores.

[1] All experimental soil pH measurements throughout this thesis were in water, i.e. $pH_{(H2O)}$

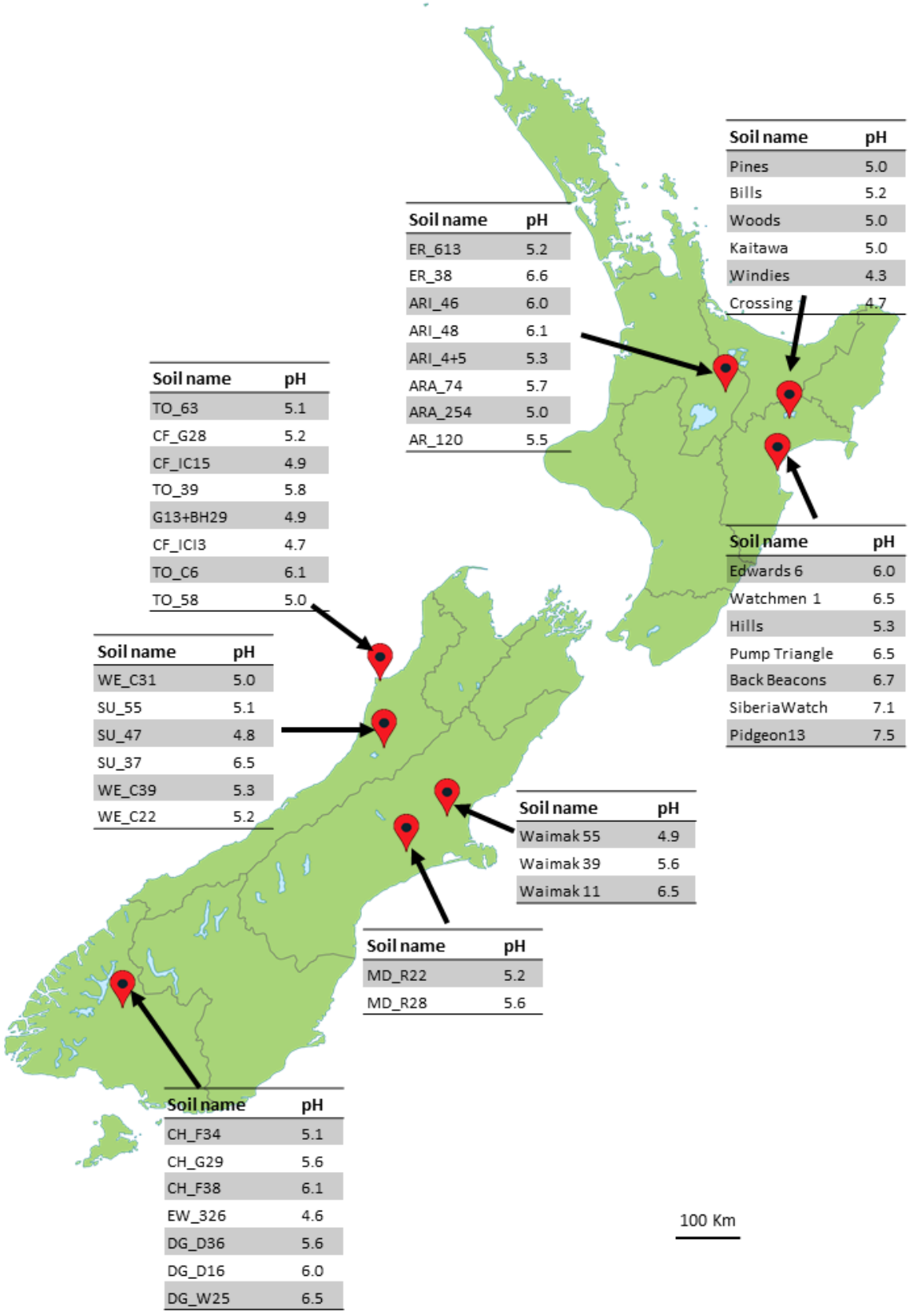

Figure 2.2: Sampling locations in the North and South Islands of New Zealand, showing soil names and pH of all soils collected at each location.

2.2.2 Soil baiting for nodule collection

Each of the 44 soils was recovered from cold storage and homogenised. Prior to sowing, seeds of white clover cv. 'Tribute' and subterranean clover cv. 'Denmark' were surface sterilised by soaking them in 70% ethanol for 1 min, followed by a 30 s soak in 4% sodium hypochlorite ($NaHClO_4$) (w/v) before rinsing six times in sterile millipore H_2O for 1 min each time. Each pottle (3 replicates per soil) contained 300 g of soil mixed with 50 g of sterile (autoclaved) coarse pumice (Intelligro Ltd., Christchurch), in which six seeds were sown, and later thinned to three seedlings upon confirmation of germination.

SC was sown on 26 May 2017 and WC on 30 May 2017. The clovers were harvested on the 24 July 2017 and 31 July 2017, respectively. The average maximum recorded temperatures in the shadehouse for June and July were 10.5°C and 8.9°C, respectively. The average minimum recorded temperatures in the shadehouse for June and July were 1.76°C and 1.55°C, respectively. The pottles were placed in large trays in a completely randomised design for eight and nine weeks for SC and WC plants, respectively. At harvest, the number of plants in each pottle was counted, shoots were cut off and bagged and roots were washed to remove excess soil. The shoots were oven-dried at 60°C for 48 h and then weighed. Shoot dry weight (SDW) for both clovers was calculated as an average of the number of surviving plants per pottle.

Roots and nodules were sterilised by soaking in 70% ethanol for 1 min, followed by soaking for 30 s in 1% $NaHClO_4$ (w/v) and then six rinses in sterile millipore H_2O for 1 min each. The roots were then stored in 20% glycerol at -20°C until the nodules could be processed. The nodules were counted for each root sample for both SC and WC, and where possible a maximum of 40 nodules per plant were collected. The target was to collect up to 120 nodules per pottle, but in most cases this number was lower, and two soils out of the 44 did not yield any nodules for either of the two clover species. This process resulted in 92 SC samples (from 36 soils) and 110 WC samples (from 41 soils), each with different numbers of nodules (Table A 12), and a total of 5,299 nodules. Pearson's correlations were assessed for the SDWs and nodule counts against soil physicochemical properties for both SC and WC.

The soil samples were split into three soil pH groups for analyses. These groups were defined both by published data on the optimal pH for clover growth and literature on the pH range of soils found in New Zealand ryegrass/clover pastures. These were: Group A (pH 4.3 – 5.3), Group B (pH 5.5 – 6.1) and Group C (pH 6.5 – 7.5). Soils in pH group A are considered suboptimal for clover growth (Howieson and Ewing 1984) because the low pH increases availability of toxic metal ions such as Al^{3+} (Unkovich et al. 1996). Thus, the suboptimal nature of this pH may drive either changes in the population of the free living rhizobia and/or specific recruitment of taxa to the nodule to alleviate stress. Although soils with pH 5.5 – 7.5 are not detrimental to clover growth this range was split into two groups. Soils in group B represented soil pH levels that are common in New Zealand pastoral farming, are moderately acidic and not detrimental to plant growth.

Neither plant nor free living rhizobia are likely to be under stress within this pH range (Moir et al. 2016). Soils of pH 6.5–7.5 (group C) are uncommon in New Zealand farming systems and soils of pH >7.0 are particularly rare. It is unknown whether soils with a pH above 6.5 could drive changes in the population of free living rhizobia or the recruitment of different taxa to the nodules. The placement of these soils into Group C allowed this possibility to be explored.

2.2.3 DNA extraction, PCR, product purification and sequencing

The nodules represented the population of rhizobia and other microbial taxa recruited by the plant from each of the soils. The nodules from each of the 92 SC and 110 WC samples were placed in 96-well plates (2 mL deep) with sterile PCR grade water and macerated using sterile glass rods to crush the nodules. A MoBio PowerPlant® Pro-htp 96-well DNA isolation kit was used to extract DNA from the samples (protocol as per kit instructions).

The eluted DNA was stored at -20°C until amplification by PCR. A total of 34 barcoded primer combinations for each of the *16S rRNA* (V3-V4 hypervariable region) (Vasileiadis et al. 2012) and *nod*C genes were used (Appendix 2: Supplementary sequencing data). For each barcoded primer combination, a negative control was also included for every PCR run (template substituted with H_2O). A 50 µL reaction was used: Kapa HiFi HotStart Readymix 25 µL, H_2O 20 µL, forward primer (10 µM) 1.5 µL, reverse primer (10 µM) 1.5 µL, DNA template (approx. 2 to 50 ng/µL) 2 µL. For PCR cycle details refer to Appendix 2. To confirm presence of the specific *16S rRNA* bands and to check the absence of bands in the negative controls, 4 µL of each *16S rRNA* PCR products was on and separated a 1% agarose gel by electrophoresis at 100 V for 45 min. After staining with ethidium bromide solution (5×10^{-4} gL^{-1}) for 15 min and de-staining in water for 5 min, bands were viewed under UV light using the VersaDoc™ Imaging System (Model 3000, Bio-Rad, CA, USA). The PCR products were purified using the QIAquick® PCR product purification kit (Qiagen). For each of the *nodC* PCR products, 45 µL of the product was loaded onto a 1% agarose gel in 1× TAE buffer (Appendix 1: Media and reagents) and separated by electrophoresis at 100 V for 45 min. After staining with ethidium bromide as previously described, bands were viewed under UV light and the specific band corresponding to the *nodC*-amplified product was excised using a sterile scalpel and stored in a 1.7 mL tube at -20°C. The bands were purified using the Zymoclean™ gel DNA recovery kit (Zymoresearch). Purified *16S rRNA* and *nodC* products were quantified using the Qubit™ dsDNA High Sensitivity assay kit (ThermoFisher Scientific). The assay is accurate for sample concentrations from 10 pg/µL to 100 ng/µL.

Samples were sent for sequencing to Novogene Company Limited (Hong Kong). Novogene undertook the quality control on the DNA (Appendix 2), library preparation – including a five-step PCR, checking the library, sequencing PCR products using HiSeq 2500 (Illumina) and filtering of raw data.

2.2.3.1 Bioinformatics and Statistical analysis

From the raw sequencing data, paired-end reads were identified and operational taxonomic units (OTUs) assigned by Aurélie Laugraud (Bioinformatician, AgResearch Ltd, Lincoln). Reads were demultiplexed and trimmed, using Flexbar v3.0 (Dodt et al. 2012; Roehr et al. 2017). The OTU analysis was done with QIIME 2™ against Greengenes database v13.8 (http://greengenes.lbl.gov) (for *16S rRNA* sequences) and against a custom-made database of 62 sequences from NCBI BLAST for *nodC* sequences (Appendix 2: Supplementary sequencing data). De-novo OTU picking was performed using UCLUST at 97% similarity for *16S rRNA* and at 100% identity for *nodC*. The OTUs from *16S rRNA* and *nodC* sequencing were filtered to remove plant DNA, singletons and low reads (Sections 2.3.1.1 and 2.3.2.1).

Diversity indices applied to microbiota data consist of differing weights of two components: richness and evenness (Jost 2006; Wagner et al. 2018). Alpha diversity is the diversity in a single ecosystem or sample. The Shannon (H') index was used to measure alpha diversity because it provided equal weighting to richness and evenness (Wagner et al. 2018). To test whether the alpha diversity differed significantly between clover species and pH groups, a non-parametric test, the Wilcoxon rank-sum test (Mann-Whitney) was used. This test reports the pairwise adjusted (Holm) p-values. ANOVA was used to check for differences in the relative abundance of each of the non-*Rhizobium* genera between the clover species.

Beta diversity refers to the diversity between samples. This is a measure of how similar or dissimilar the samples are, and is usually represented by a distance matrix which is then used to do Principal Coordinates Analysis (PCoA). The Bray-Curtis dissimilarity metric was used as a measure of beta diversity for *16S rRNA* and this was derived from the OTU frequencies (relative abundances) between samples.

The *nodC* gene was used to explore diversity of rhizobia in the nodules with the aim of determining whether there was a relationship between genotype and soil pH. NodC is a functional protein involved in nodulation, thus, even a single nucleotide substitution (SNP) in a key codon could alter function. This is unlike genes used for the purposes of taxa identification where groupings into OTUs at 97% provide sufficient resolution of diversity. Given that *nodC* amplifies from a single group of closely related taxa and that SNPs may significantly alter function, the reads were placed into groups of 100% amino acid identity. For *nodC*, the sequences of the remaining OTUs (after filtering process) were translated into protein sequences in Geneious version 11.1.4 (https://www.geneious.com/; and also by Aurélie Laugraud). This resulted in 353 unique (at least one amino acid change) protein sequences. The protein sequences were aligned using MUSCLE (Edgar 2004) against the (translated) *nodC* sequences of TA1 and WSM1325 (commercial rhizobia strains). The unique protein sequences were clustered in SeaView version 4.7 (Galtier et al. 1996) by building a tree using the BIONJ algorithm with a Poisson model (Gascuel 1997) with a bootstrap value of 500. BIONJ is a phylogeny software based on the

maximum-likelihood principle. It is well suited for estimating evolutionary distances and is well adapted for estimates obtained from aligned sequences (Gascuel 1997).

Pearson's correlation analyses, alpha diversity analysis, ANOVA and graphing boxplots of *16S rRNA* and *nodC* data were performed in R (v 3.4.3) (Team 2017) using the platform R Studio (v 1.1.419). The 'ggplot2' package was used for graphing (Wickham 2011) and the 'phyloseq' package was used for alpha diversity analysis (McMurdie and Holmes 2013). Bray-Curtis resemblance matrices and PCoA's were performed using PRIMER v7 (Clarke and Gorley 2015). Permutation based multivariate analysis of variation (PERMANOVA; (Anderson et al. 2008)) was used to determine if relative abundances of non-*Rhizobium* taxa were related to soil pH and/or sampling region (999 permutations). PERMANOVA is one of most widely used non-parametric methods to fit multivariate models to microbiome data. It is a multivariate analysis of variance based on distance matrices and permutation (McArdle and Anderson 2001; Xia and Sun 2017). PERMANOVAs were performed on square-root transformed abundance data for each sample, from which the beta-diversity (dissimilarity) among samples was calculated using the Bray-Curtis method. The number of reads (relative abundance) for the representative *nodC* OTUs were analysed by running Generalised Linear Models (GLMs) using the Quasi Poisson distribution in R.

2.3 Results

2.3.1 Analysis of nodule microbiome using *16S rRNA* data

2.3.1.1 Sequencing information

A total of 20,563,699 raw sequences were returned for the *16S rRNA* gene. From these, 18,616,143 paired-end amplicon reads were identified and grouped into 1,219,953 OTUs based on matches with 97% similarity in the Greengenes database. Filtering of OTUs which were designated as chloroplast (~6%; 70,887), mitochondria (~10%; 123,153) and unassigned (~4%; 42,797) removed ~21% of the paired-end amplicon reads. The remaining 14,803,307 paired-end reads were placed into 983,116 OTUs. Finally, after filtering out OTUs with less than 1,000 reads (Lundberg et al. 2012), 283 OTUs remained with 12,609,804 paired-end amplicon reads, across the 202 samples. Of the 283 OTUs, 134 were assigned to the genus *Rhizobium* (with 11,410,746 paired-end reads) and 149 to 'non-*Rhizobium*' genera (with 1,199,058 paired-end reads). From the 'non-*Rhizobium*' OTUs, 67 were not identified at the genus level and were labelled as 'Undefined'. 'Undefined' was categorised into at least eight families (Table A 8) and accounted for 4.4% of total reads. In total, 34 genera were identified (Appendix 2: Supplementary sequencing data). The three most abundant OTUs (assigned to genus *Rhizobium*) made up 10,925,054 reads, ~87% of the total.

The average number of OTUs and reads for each pH group for each clover species are shown in Table A 9 (Chapter 2 supplementary results).

2.3.1.2 Alpha diversity of the complete nodule microbiome

The Shannon index was used to compare alpha diversity between SC and WC samples and also between pH groups of each clover species. Overall, the alpha diversity was higher ($p < 0.0001$) in SC samples than the WC samples, H' index of 1.659 (SC) versus 1.246 (WC). There was no significant difference in H' index between pH groups for both clover species, although there was a strong trend ($p = 0.058$) for a difference between groups A and B for SC samples (Tables 2.1 and 2.2).

Table 2.1: Average Shannon index values for total nodule microbiome in pH groups A, B and C for each of SC and WC samples.

	Shannon (H') index		
pH group	A (pH 4.3 – 5.3)	B (pH 5.5 – 6.1)	C (pH 6.5 – 7.5)
Subterranean Clover (SC)	1.7991	1.4914	1.5898
White Clover (WC)	1.2577	1.2226	1.2493

Table 2.2: Pairwise comparisons of Shannon index for total nodule microbiome using Wilcoxon rank sum test, showing holm-adjusted p-values.

	Subterranean Clover		White Clover	
pH group	**A**	**B**	**A**	**B**
B	0.058	-	0.272	-
C	0.499	0.417	0.320	0.991

2.3.1.3 Comparative alpha diversity of Rhizobium and non-Rhizobium genera

It was expected to find an abundance of *Rhizobium* in the nodules of both clover species, compared with other bacterial genera, and this was the bulk of the read data. *Rhizobium* species have a known function in nodules, whereas, the functional significance of other taxa is not known. Thus, to better explore the effect of pH, these two distinct groups (*Rhizobium* and "other taxa") were analysed separately.

Rhizobium diversity

Relative to all the genera present in the nodules, *Rhizobium* had an abundance of 82.7% in SC nodules and 96.7% in WC nodules. The Wilcoxon rank sum test was used to compare the alpha diversity of the genus *Rhizobium* between SC and WC, and showed a significant difference ($p < 0.0001$) between these clover species, with the alpha diversity of SC (H' = 1.206) higher than that of WC (H' = 0.966). However, there was no difference in alpha diversity (H') of rhizobia between pH groups for either clover species, although there was a trend ($p = 0.085$) for differences between groups A and B in SC samples (Table 2.3), with the alpha diversity of group A (H' = 1.292) higher than that of group B (H' = 1.111).

Table 2.3: Pairwise comparisons of Shannon index for genus *Rhizobium* using Wilcoxon rank sum test, showing holm-adjusted p-values.

	Subterranean Clover		White Clover	
pH group	**A**	**B**	**A**	**B**
B	0.085	-	0.212	-
C	0.532	0.532	0.890	0.890

Non-Rhizobium diversity

The Wilcoxon rank sum test was used to compare the alpha diversity of non-*Rhizobium* genera between SC and WC, and showed a significant difference in H' index ($p < 0.0026$) between the clover species, with the alpha diversity of WC (H' = 3.435) higher than that of SC (H' = 3.200). There were also differences between pH groups A and B for both SC and WC, ($p = 0.031$ and 0.026, respectively) and groups B and C ($p = 0.031$) for SC (Tables 2.4 and 2.5). There was also a strong trend for differences between pH groups B and C ($p = 0.067$) for WC.

Table 2.4: Average Shannon (H') index values for non-*Rhizobium* taxa in pH groups A, B and C for each of SC and WC samples.

	H' index (non-*Rhizobium*)		
pH group	A (pH 4.3 – 5.3)	B (pH 5.5 – 6.1)	C (pH 6.5 – 7.5)
Subterranean Clover (SC)	3.082	3.333	3.272
White Clover (WC)	3.445	3.418	3.434

Table 2.5: Pairwise comparisons of Shannon index for non-*Rhizobium* genera using Wilcoxon rank sum test, showing holm-adjusted p-values. Values with * are significant at 95%.

	Subterranean Clover		**White Clover**	
pH group	**A**	**B**	**A**	**B**
B	0.031*	-	0.026*	-
C	0.751	0.031*	0.977	0.067

Since there was a significant difference in alpha diversity of nodule bacteria between SC and WC ($p < 0.0001$), it was necessary to elucidate the sources of the difference. Table 2.6 shows the relative abundance of different genera in each clover species. There were no unique taxa in either clover, but the relative proportions of taxa in the nodule biome differed between the hosts. SC had significantly higher relative abundance of 18 of the 33 non-*Rhizobium* genera compared with WC (Table 2.6). None of the genera had significantly higher relative abundance in WC compared with SC.

Table 2.6: Distribution of reads from OTUs of bacterial genera from *16S rRNA* sequences for each genus in the SC and WC samples. Genera with relative abundance significantly ($\alpha = 0.05$) higher in SC samples compared with WC samples, are highlighted in **bold red**. SC = Subterranean clover, WC = White clover.

Genus	Relative abundance % (SC)	Relative abundance % (WC)	Total reads	Total OTUs
Pseudomonas	**3.754**	1.246	297063	15
Sinorhizobium	**1.844**	0.118	111218	3
Salmonella	**1.174**	0.090	71848	1
Delftia	0.292	0.029	18283	2
Bacillus	**0.231**	0.045	16050	5
Novosphingobium	0.093	0.054	8991	4
Janthinobacterium	0.094	0.045	8447	4
Burkholderia	**0.120**	0.024	8389	5
Erwinia	0.034	0.084	7833	3
Agrobacterium	0.058	0.064	7717	2
Bradyrhizobium	**0.093**	0.028	7195	2
Sporosarcina	**0.104**	0.019	7128	2
Propionibacterium	**0.093**	0.018	6463	2
Streptomyces	0.061	0.040	6163	2
Devosia	**0.074**	0.024	5850	2
Thalassospira	0.009	0.075	5816	2
Sphingomonas	**0.074**	0.017	5343	2
Caulobacter	0.067	0.013	4652	2
Ochrobactrum	**0.066**	0.010	4420	2
Asteroleplasma	**0.065**	0.007	4150	2
Stenotrophomonas	**0.069**	0.002	4007	2
Flavobacterium	0.036	0.012	2854	2
Lysinibacillus	**0.042**	0.007	2847	2
Rhodanobacter	0.033	0.014	2839	1
Bacteroides	**0.049**	0.000	2753	2
Asticcacaulis	0.017	0.025	2744	1
Anaerobacillus	**0.032**	0.005	2104	2
Acidovorax	0.010	0.013	1433	1
Mycoplana	0.004	0.014	1223	1
Phenylobacterium	0.008	0.008	1036	1
Rahnella	**0.015**	0.003	1033	1
Rhodoferax	**0.016**	0.002	1021	1
Inquilinus	0.004	0.011	1019	1

The effect of pH groups on the relative abundance of individual taxa is shown in Figure 2.3, with *Pseudomonas*, *Salmonella* and *Sinorhizobium* being the genera (highlighted) for which there was the most variance between pH groups.

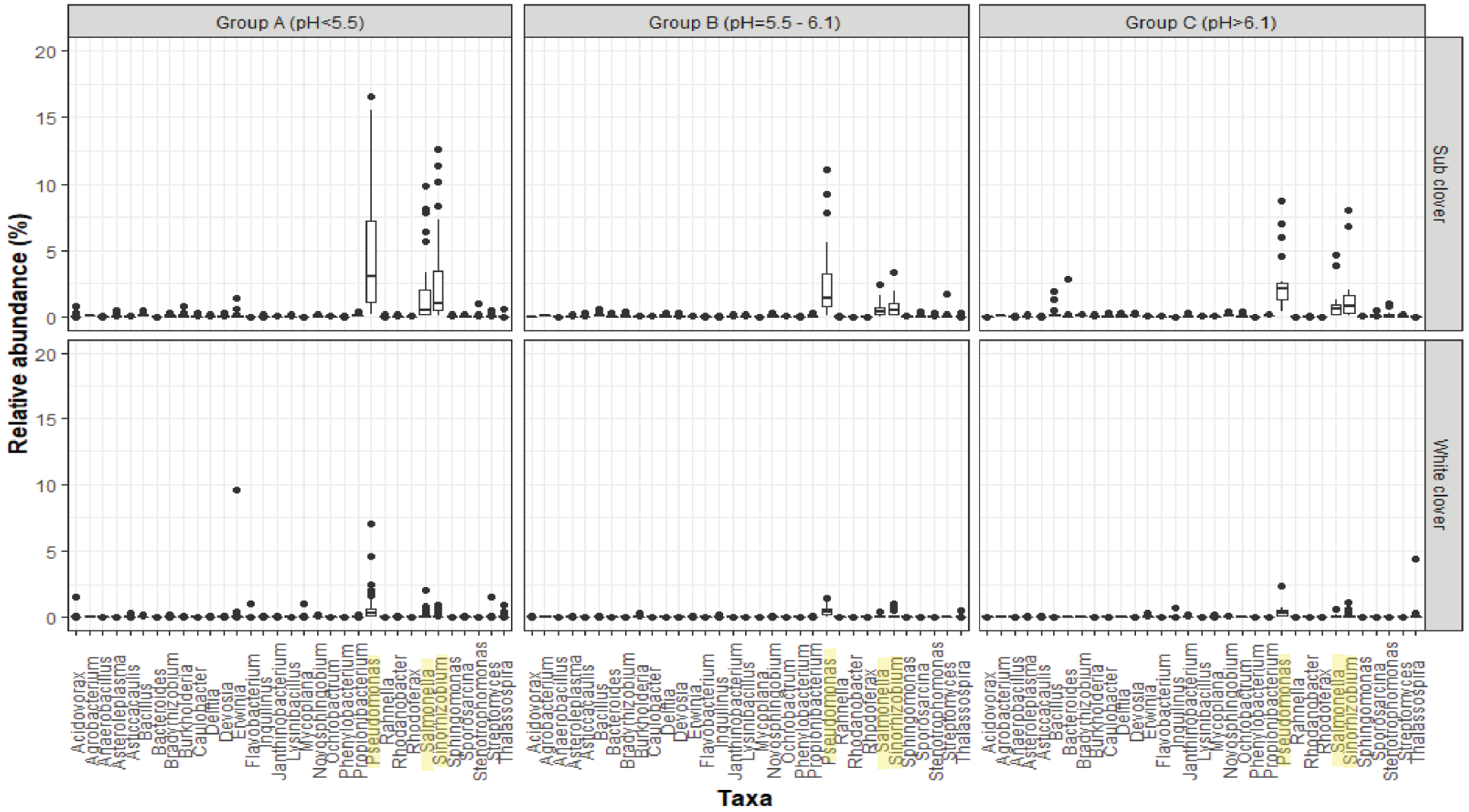

Figure 2.3: Boxplot of the relative abundance of 'non-*Rhizobium*' genera for SC and WC samples for each of the three pH groups. SC = Subterranean clover, WC = White clover.

2.3.1.4 Beta diversity of Rhizobium and non-Rhizobium genera

The similarity in communities of *Rhizobium* and non-*Rhizobium* bacteria present on WC and SC nodules were compared using PCO ordination. For *Rhizobium*, a total of 83.9% of the variability was explained using only two PCOs (Figure 2.4), and 65.8% of the variability was explained for the non-*Rhizobium* genera (Figure 2.5). The clover species explained all of this variability in beta diversity for both *Rhizobium* and non-*Rhizobium* genera.

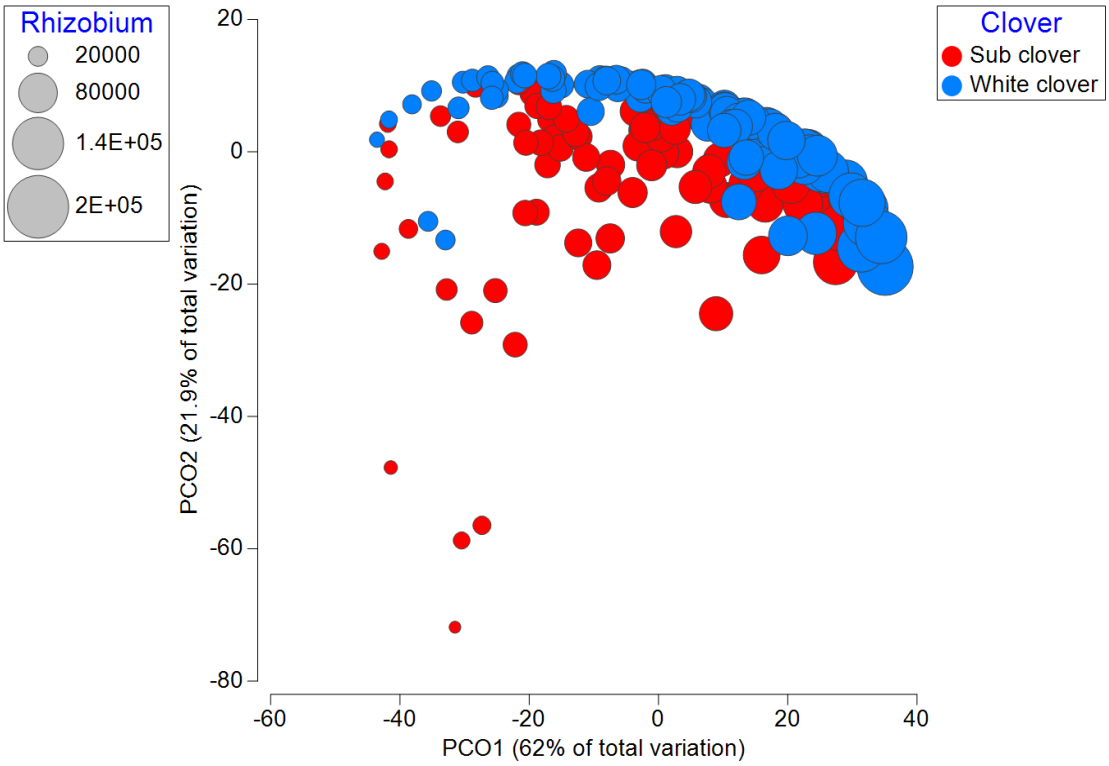

Figure 2.4: A bubble plot generated from PCO analysis of Bray-Curtis resemblance data for only genus *Rhizobium*. The colours differentiate the samples according to clover species. The sizes of the circles correspond to the number of reads of *Rhizobium*.

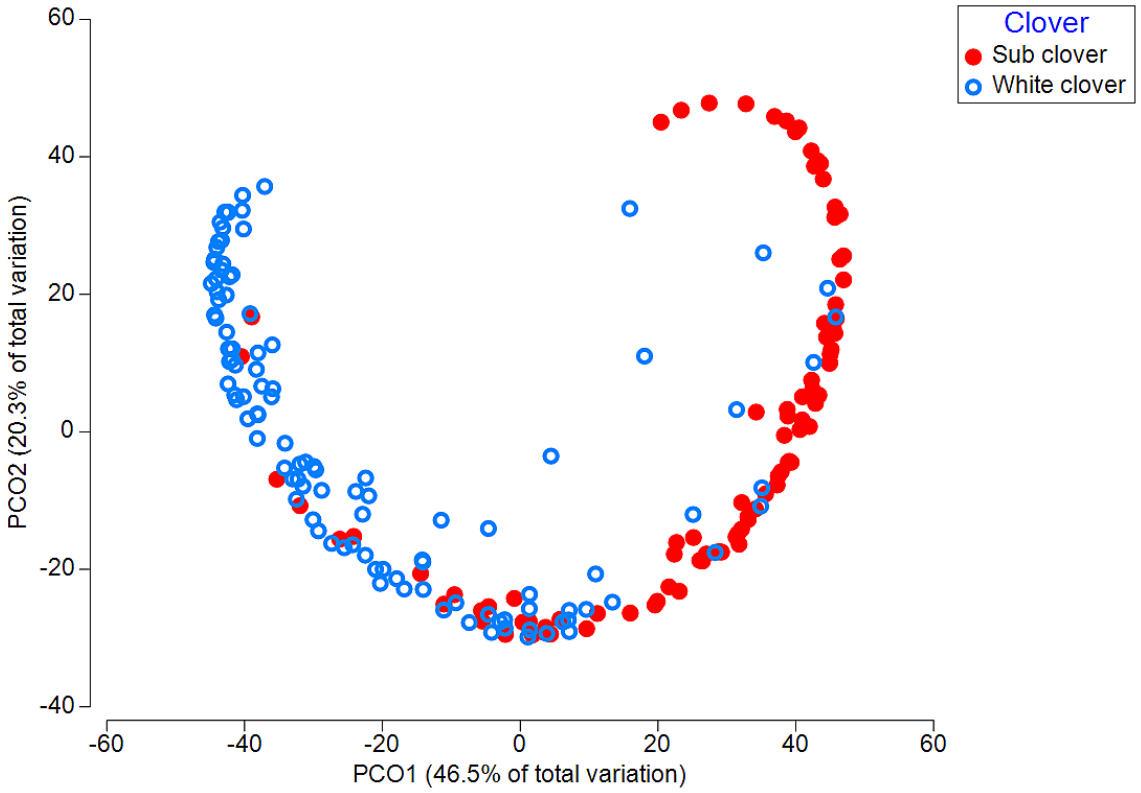

Figure 2.5: PCO analysis of Bray-Curtis resemblance data from the samples for all non-*Rhizobium* genera. The symbols differentiate the samples according to the clover species.

2.3.1.5 Correlations of alpha diversity and relative abundance with soil properties

As the alpha diversity was significantly different ($p < 0.0001$) between SC and WC, they were analysed for correlations with soil physicochemical factors separately.

Table A 10 (Chapter 2 supplementary results) shows the Pearson's correlations separately for the relative abundance and H' index of *Rhizobium* and non-*Rhizobium* genera with soil properties (other than pH) for SC and WC samples. The p values show significance at 95% confidence intervals. *Rhizobium* diversity in SC nodules was positively associated with Al ($r = 0.2066$, $p = 0.0481$). For WC samples, Olsen P ($r = 0.2433$, $p = 0.0104$) and K ($r = 0.1942$, $p = 0.0420$) were positively associated with *Rhizobium* abundance in WC nodules. The diversity of non-*Rhizobium* bacteria in WC samples was positively associated with Olsen P ($r = 0.2111$, $p = 0.0269$) and Mg ($r = 0.1945$, $p = 0.0417$).

2.3.2 *nodC* data

2.3.2.1 Sequencing information

A total of 18,996,526 raw sequences were returned for the *nodC* gene. These sequences were compared with 62 *nodC* sequences retrieved from the NCBI database that had been obtained by BLAST search (Appendix 2: Supplementary sequencing data). The sequences spanned 8 genera: *Bradyrhizobium*, *Burkholderia*, *Mesorhizobium*, *Microvirga*, *Neorhizobium*, *Phyllobacterium*, *Rhizobium* and *Sinorhizobium*. This reference database was curated to assign taxonomy to the *nodC* OTUs. Based on 100% similarity with an OTU, 20,083,434 paired-end amplicon reads were identified and grouped into 1,004,502 OTUs. There were 27,633 OTUs with no BLAST hits (~2.8% of total), making up 107,791 reads (~0.5% of total) and these were removed. Further filtering excluded 976,069 OTUs (~97% of total) with fewer than 1,000 reads from the analysis. The remaining 16,672,670 paired-end reads (~83% of total) were placed into 800 OTUs across the samples. There were 594 unique DNA sequences from the 800 OTUs which, when translated into protein sequences, resulted in 353 unique amino acid sequences (OTUs) with a total of 944,252 reads across the samples. Of these, 516,915 reads were from the 92 SC samples (54.7%) and 427,337 reads were from the 110 WC samples (45.3%). The 353 OTUs were most similar to four taxa (from the database): *Rhizobium leguminosarum* strain R460.2 (252 OTUs with 707,105 reads), *Rhizobium* spp. CCBAU 83304 (87 OTUs with 194,058 reads), *R. leguminosarum* strain R45915 (13 OTUs with 41,013 reads) and *R. bangladeshense* strain 1017 (1 OTU with 2076 reads).

2.3.2.2 Clustering OTUs

A BIONJ tree was built using the protein sequences of the 353 OTUs. Two major groups were distinguished in the radial form of the tree (Figure 2.6), with Group 1 containing a majority (73%) of the OTUs. The sequences of the commercial strains clustered separately, with one in each of the two groups; TA1 in Group 1 and WSM1325 in Group 2. The OTUs in Group 1 were most similar to *R. leguminosarum* strain R460.2, whereas the ones in Group 2 were most similar to either *Rhizobium* spp. CCBAU 83304 or *R. leguminosarum* strain R45915. From each group, eight representative OTUs were selected to assess their association with either clover species and/or pH group. These OTUs were selected from collapsed nodes and thus, were representatives of genotypes (genetically similar strains) within the major branches.

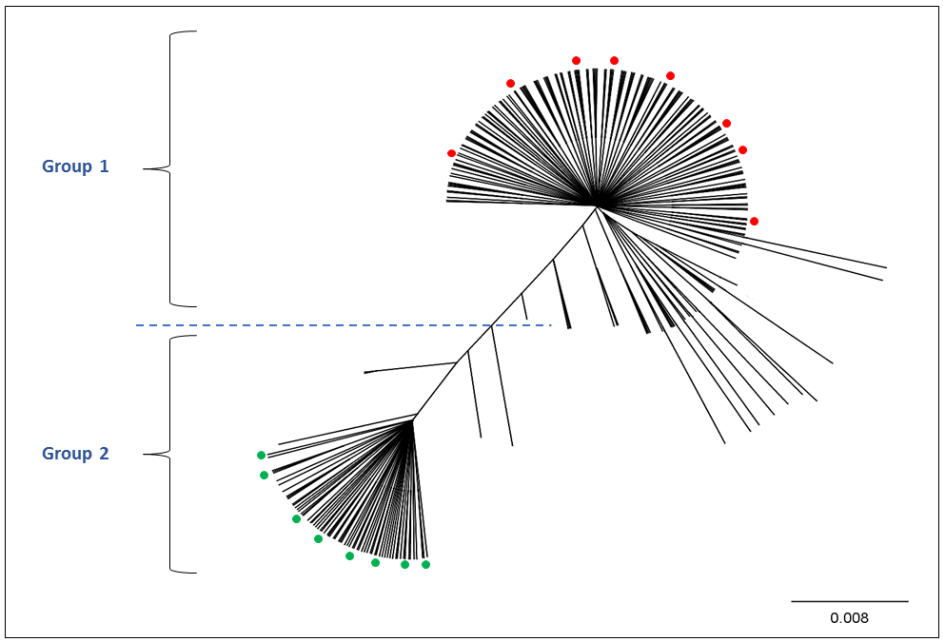

Figure 2.6: BIONJ tree visualised in a radial format, showing the two major groups in which the NodC OTUs cluster. The representative OTUs from each group are marked with either • or •.

Table 2.7 shows abundance data of the eight representative NodC OTUs from each major group of the BIONJ tree (Figure 2.6). The values are expressed as percentage reads (frequency) in each clover species and split between two pH groups (selected based on the results from the *16S rRNA* data).

The GLM analysis for the representative OTUs is presented in Chapter 2 supplementary results. The results showed that some NodC genotypes occurred more frequently within a particular host species or soil pH group. OTU 8 occurred at higher frequency in SC plants grown in soils from Group C, whereas OTUs 2 and 10 were more frequently recovered from SC grown in soils from Group A. OTU 6 occurred more frequently in WC samples from Group A soils. OTU 17 was more frequent in plants grown in soils from Group C (no difference between clover species), and OTU 15 was more commonly recovered from WC (no difference between pH groups).

Table 2.7: Relative abundance (% reads) of representative *nodC* OTUs from the two major groups, in both clovers and in pH <5.5 or pH >5.5. Highlighted OTUs have significant differences ($p \leq 0.05$) in frequency between either clovers or pH groups. SC = Subterranean clover, WC = White clover.

		SC		WC		Taxonomy assignment
	OTU ID	**pH < 5.5**	**pH > 5.5**	**pH < 5.5**	**pH > 5.5**	
Group 1 (includes TA1)	4.	27.65	28.16	22.19	22.00	*R. leguminosarum* Strain 460.2
	1.	19.65	35.62	21.32	23.41	*R. leguminosarum* Strain 460.2
	8.	24.80	43.74	13.17	18.29	*R. leguminosarum* Strain 460.2
	5.	26.07	31.50	22.34	20.09	*R. leguminosarum* Strain 460.2
	7.	22.58	27.93	23.32	26.18	R. leguminosarum Strain 460.2
	6.	15.37	6.39	45.25	32.99	*R. leguminosarum* Strain 460.2
	2.	67.17	4.01	27.78	1.04	*R. leguminosarum* Strain 460.2
	3.	23.53	27.73	27.82	20.92	*R. leguminosarum* Strain 460.2
Group 2 (includes WSM1325)	17.	3.63	31.99	13.66	50.72	*R. leguminosarum* Strain 45915
	10.	94.24	1.16	3.43	1.17	*R. leguminosarum* Strain 45915
	12.	10.44	17.86	13.46	58.24	*R. leguminosarum* Strain 45915
	14.	17.30	28.55	27.35	26.81	*Rhizobium* spp. CCBAU 88304
	13.	9.29	16.21	22.00	52.50	*Rhizobium* spp. CCBAU 88304
	15.	11.51	14.58	44.31	29.60	*Rhizobium* spp. CCBAU 88304
	16.	15.30	29.69	26.72	28.29	*Rhizobium* spp. CCBAU 88304
	11.	16.45	27.8	29.51	26.25	*Rhizobium* spp. CCBAU 88304

Figure 2.7 shows the amino acid alignment of the 16 representative OTUs as well as the commercial strains TA1 and WSM1325. The colours of the amino acids are based on their polarities and showcase small differences in the alignment, i.e. at sites 30, 44, 67, 79, 92, 102 and 104. The two motifs (DX and RW) and the transmembrane domain are allocated as per the alignment of the *Sinorhizobium meliloti* NodC protein (Dorfmueller et al. 2014).

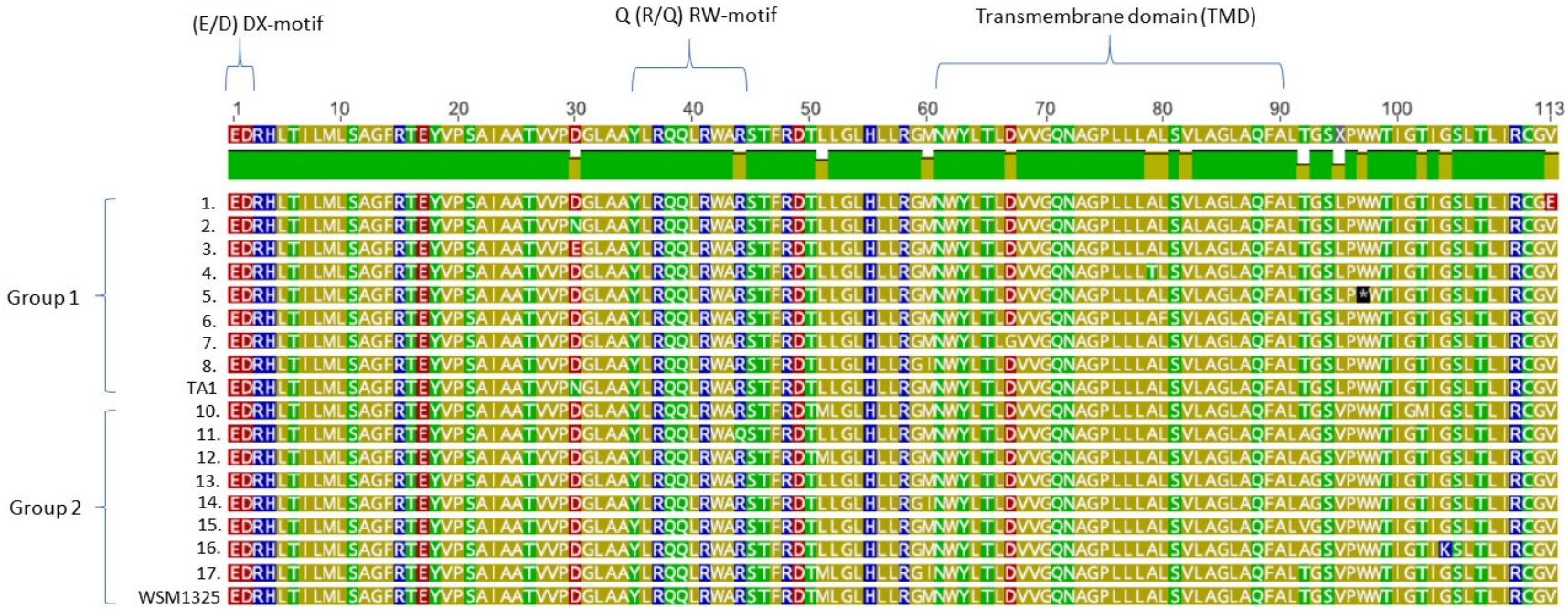

Figure 2.7: Alignment of the translation of representative OTUs from each group including the commercial strains, TA1 and WSM1325. The figure has been annotated with information about the NodC protein from Dorfmueller *et al.* (2014).

2.3.2.3 nodC alpha diversity

The Shannon index was used to measure NodC alpha diversity between clover species and between pH groups of each species. The Wilcoxon rank sum test showed no significant differences in alpha diversity of NodC between clover species or between pH groups within each clover species.

2.3.3 Soil physicochemical properties

Soil physicochemical properties are often not independent, and there are known relationships between pH and mineral content, or soil type. The Pearson's correlations of pH with other soil properties of all soils are shown in Table 2.8, and the graphical depiction is shown in Figure A 1 (Chapter 2 supplementary results). Soil properties in blue have significant positive correlations, whereas those in red have significant negative correlations with pH. The measurements of minerals were derived from extractable fractions, i.e. ammonium acetate extraction for K, Ca, and Mg and calcium chloride extraction for Al.

Table 2.8: Pearson's correlations of soil pH with other measured soil physicochemical parameters, and their corresponding p values. Values with * are significant at 95%.

	Soil pH	
	Pearson's r	p value
Olsen Phosphorus	-0.0426	0.7785
Potassium	0.3608	0.0138 *
Calcium	0.7610	< 0.0001 *
Magnesium	0.3742	0.0104 *
Aluminium	-0.6637	< 0.0001 *
Organic Matter	-0.4651	0.0011 *

2.3.4 Shoot Dry Weight and Nodule data

The shoot dry weights (SDWs) were measured for both clover species, from all soils which yielded live plants and the nodules were counted. Pearson's correlations for the SDWs and nodule counts against soil physicochemical properties for SC and WC plants are presented in Table 2.9 and Table 2.10, respectively.

SDWs for SC had positive correlations with Olsen P ($p < 0.0001$), K ($p = 0.0001$) and OM ($p = 0.0456$). Nodule counts for SC had positive correlations with soil pH ($p < 0.0001$) and Ca ($p = 0.0005$), and negative correlations with Al ($p = 0.0004$) and OM ($p = 0.0003$).

For WC, SDWs had positive correlations with Olsen P ($p < 0.0001$), K ($p = 0.0001$) and Ca ($p = 0.0019$). Nodule counts for WC had positive correlations with soil pH ($p = 0.0009$) and Ca ($p = 0.0244$), and negative correlations with Al ($p = 0.0005$) and OM ($p = 0.0001$).

Table 2.9: Pearson's correlations of shoot dry weight (SDW) and nodule counts for SC with measured soil physicochemical parameters, and their corresponding p values. Values with * are significant at 95%.

	Subterranean Clover (SC)			
	SDW (mg/plant)		**Nodule count**	
	Pearson's r	p value	Pearson's r	p value
Soil pH	-0.0383	0.6703	0.3786	< 0.0001 *
Olsen Phosphorus	0.4284	< 0.0001 *	0.0055	0.9510
Potassium	0.3315	0.0001 *	0.1326	0.1387
Calcium	0.0517	0.5655	0.3039	0.0005 *
Magnesium	-0.1132	0.2068	0.1122	0.2111
Aluminium	-0.0943	0.2938	-0.3108	0.0004 *
Organic Matter	0.1784	0.0456 *	-0.3146	0.0003 *

Table 2.10: Pearson's correlations of shoot dry weight (SDW) and nodule counts for WC with measured soil physicochemical parameters, and their corresponding p values. Values with * are significant at 95%.

	White Clover (WC)			
	SDW (mg/plant)		**Nodule count**	
	Pearson's r	p value	Pearson's r	p value
Soil pH	0.0694	0.4400	0.2915	0.0009 *
Olsen Phosphorus	0.3776	< 0.0001 *	0.0383	0.6705
Potassium	0.3327	0.0001 *	0.1264	0.1583
Calcium	0.2744	0.0019 *	0.2005	0.0244 *
Magnesium	0.0215	0.8112	-0.0647	0.4717
Aluminium	-0.1598	0.0738	-0.3051	0.0005 *
Organic Matter	-0.1260	0.1598	-0.3333	0.0001 *

The averages for SDW and nodule numbers for each pH group of each clover species were calculated as per the number of soils in each pH group (Table 2.11) and analysed for differences. As confirmed by Tables 2.9 and 2.10, there were no differences in SDW across the pH groups for both clover species, i.e. pH did not affect SDW. However, there were differences in nodule numbers (Table 2.12) between the pH groups. Nodule numbers between pH groups A and C ($p = 0.0001$), and groups B and C ($p = 0.024$) for SC were significantly different, whereas there was a trend between groups A and C ($p = 0.072$) for WC.

Table 2.11: Average SDW and nodule number data for each pH group of each clover species.

	Subterranean Clover		**White Clover**	
	SDW (mg/plant)	Nodules	SDW (mg/plant)	Nodules
Group A (pH < 5.5)	126.76	45.08	76.83	57.92
Group B (pH 5.5–6.1)	129.41	58.42	78.71	67
Group C (pH > 6.1)	121.04	86.25	85.80	79

Table 2.12: Holm-adjusted p-values from pairwise comparisons of nodule numbers between pH groups.

	Subterranean Clover		**White Clover**	
pH group	**A**	**B**	**A**	**B**
B	0.1175		0.491	
C	0.0001*	0.024*	0.072	0.491

2.4 Discussion

2.4.1 Summary of findings

The goal of this chapter was to determine whether there was a relationship between soil pH and the diversity of bacteria inhabiting the nodules of SC and WC, as well as to compare between the nodule community structure of SC and WC. This work contributed to the increasing understanding of how plants modify their microbiome in response to abiotic stress. Although research has been done on plant roots, the nodule is a specialised structure with a primary role of N fixation and it was unknown whether the external soil pH would affect the microbial communities within the nodule. To achieve this goal, the bacterial communities were assessed from 5,299 nodules of SC and WC grown in 44 soils that varied in their pH and edaphic properties.

The main outcome of this chapter was that there was higher alpha diversity of the total nodule microbiome in SC nodules compared with WC nodules. There was a strong trend supporting higher diversity in pH group A compared with group B in SC nodules. Clover species was a stronger driver of beta diversity of both *Rhizobium* and non-*Rhizobium* genera than soil physicochemical properties. Irrespective of pH, the SC cultivar harboured higher relative abundance of non-*Rhizobium* bacteria in nodules compared with the WC cultivar. The results showed that, as expected, at low soil pH fewer nodules were formed on both clovers. There was no overall effect of soil pH or other soil properties on the diversity and abundance of *nodC* in nodules of SC and WC. However, the *nodC* OTUs clustered in two major groups, with some genotypes demonstrating preference to host, soil pH or both.

2.4.2 *16S rRNA* data

The goal of this work was to determine whether soil pH and/or clover species could alter the recruitment of microbial taxa to the nodule, changing the nodule biome by influencing diversity and/or abundance of bacteria communities within nodules. pH has shown to be a first-order factor which influences the soil bacterial community (Lauber et al. 2009), but it's effect on the nodule microbiome is unknown. The expectation was that both low and high soil pH would cause a decrease in diversity and/or abundance of nodule bacteria, and that clover species would have no or only a minor influence. This work demonstrated that the total nodule biome of the SC cultivar was more diverse than that of the WC cultivar. The SC cultivar also had a higher diversity of *Rhizobium* than the WC cultivar, but non-*Rhizobium* genera were more diverse in WC than in SC nodules. SC nodules had higher alpha diversity in the total nodule bacterial community as well as the *Rhizobium* in pH group A than in group B. This was not the case for the WC cultivar.

Rhizobia (composed of *Rhizobium*, *Bradyrhizobium* and *Sinorhizobium* genera) were dominant in the nodules from both clovers, making up ~92% of the total reads. This was expected as these

taxa initiate nodulation and was similar to the results of previous studies on legume species. From the 16S rDNA sequencing of rhizoplane-endorhizosphere soil planted with *T. repens*, Marilley and Aragno (1999) found that there was a dominance of *R. leguminosarum* and associated that with the presence of nodules on the clover roots. Another study that involved *16S rRNA* gene sequencing of 67 bacterial isolates from nodules of *T. repens*, *T. fragiferum*, *Mimosa pudica*, *M. invisa* and *Crotalaria pallida* found that the majority of the isolates belonged to *Rhizobium*, *Bradyrhizobium, Sinorhizobium, Agrobacterium* and *Mesorhizobium* genera (Liu et al. 2007). The three most abundant OTUs from these results (matching *Rhizobium* spp.) made up 87% of total reads. Other studies have shown similar proportions. Hartman *et al.* (2017) found dominance of one OTU, matching *R. leguminosarum* (which accounted for a median of 73.5%), in all their root samples of *T. pratense* (red clover). Seth's (PhD thesis, Lincoln University, 2017 – embargoed until August 2020) work also identified that ~86% of reads from the most abundant OTUs matched with *Rhizobium* spp..

The bacterial taxa other than rhizobia identified from the *16S rRNA* sequencing in this chapter are similar to previous findings by Hartman *et al.* (2017). They found OTUs representing seven different orders of bacteria: Rhizobiales, Sphingomonadales, Enterobacteriales, Burkholderiales, Caulobacterales, Rhodospirillales (all from Proteobacteria) and Firmicutes (non-Proteobacteria) from the root microbiome of *T. pratense*. They performed *16S rRNA* amplicon sequencing using the Illumina MiSeq platform, but their 16S primers flanked the V5-V7 regions rather than the hypervariable V3-V4 regions used in this present study. Previous work by Seth (2017) and John Ramana (MSc thesis, 2018) have shown similar patterns in *16S rRNA* (V3-V4 regions) sequencing of nodules to the present study. In 120 SC and WC nodules, Seth (2017) found the presence of the following non-rhizobia genera: *Sphingomonas*, *Novosphingobium*, *Bacillus*, *Pseudomonas*, *Enterobacter* and *Variovorax*. Ramana (2018) sequenced the root microbiome of 18 WC plants (from six different cultivars) and found the phyla Proteobacteria and Bacteriodetes to be the most common, followed by Actinobacteria and Firmicutes.

Beta diversity was driven by the clover species, explaining ~84% and ~67% of the variability for *Rhizobium* and non-*Rhizobium* genera, respectively, in the PCO analysis. The edaphic factors had no influence on beta diversity of the bacterial taxa (Chapter 2 supplementary results).

The research described in this chapter has produced sequence data that is similar to two recent New Zealand studies that also used Illumina amplicon sequencing to investigate bacterial populations in pasture legumes, including WC and SC. Seth (2017) evaluated nodule communities in SC and WC across low and high phosphate fertiliser regimes, and Wigley (2017) studied the comparative nodule microbiome in WC and lucerne. The 120 nodules sequenced by Seth (2017) produced 4,720,004 reads, of which ~35% were allocated to plant (mitochondrial or chloroplast) DNA. Wigley (2017) sequenced 30 WC nodules and found ~10% of reads matched plant DNA.

These percentages are consistent with the ~16% plant DNA reads removed from the present study.

2.4.2.1 Differences in nodule microbiome of subterranean and white clover

This work showed that the relative abundance of non-*Rhizobium* genera was consistently greater in SC than in WC nodules across the three pH groups (Table 2.6). The differences in relative abundance of bacterial genera between the clover species may be linked to differences in root exudates, lectins, isoflavonoids or rhizodeposits (Hirsch 1999; Berg and Smalla 2009), either due to i) passive inclusion into the nodule during nodule formation, or ii) active recruitment/selection of bacteria for their functional characteristics (Seth 2017). Dennis *et al.* (2010) reviewed the role of root exudates and/or rhizodeposists on bacterial community structure. The authors suggested that the direct influence of root exudates on rhizosphere microbial communities is limited to the small area surrounding root apices and that the role of root exudates in structuring rhizosphere bacterial communities needs to be considered in the context of the wider contribution of other rhizosphere carbon pools (Dennis et al. 2010). However, Seth (2017) found no difference in alpha or beta diversity of the rhizosphere bacterial communities between SC and WC, suggesting that root exudates from the two clovers may not be driving bacterial communities inside the nodule. Furthermore, Gomaa *et al.* (2015) identified and quantified 13 flavonoids using HPLC and found no differences in the flavonoid (aglycones or glycosides) profiles between shoots of *T. alexandrinum* and *T. resupinatum*. Other studies have confirmed that the rhizosphere bacterial community is both qualitatively and quantitatively different from root endophytic communities (Lundberg et al. 2012; Bulgarelli et al. 2015; Hartman et al. 2017).

WC nodules had greater alpha diversity of non-*Rhizobium* bacteria than SC nodules, with pH group A having the highest alpha diversity for WC. Interestingly, the alpha diversity in pH group B for SC was highest, but it was lowest in group B for WC (Table 2.4). However, of the taxa identified, no genus was completely excluded from either clover, i.e. all observed genera were present in at least some samples of both SC and WC. This is consistent with Seth's (2017) research, which showed that the overall bacterial species richness of 120 nodules of WC and SC was not different. The greater relative abundance of non-*Rhizobium* genera in SC nodules may occur because this plant species actively includes or excludes certain bacteria in order to cope with edaphic stresses or exploit specific nutrient reserves in soil. This may be mediated to some degree by differences in exudate profile, especially under a stress such as pH. The work of Seth (2017) was focussed on the identification of phosphate solubilising bacteria (PSB), which found that SC nodules contained more non-rhizobial PSB than WC nodules, indicating a functional role in bacterial recruitment into SC nodules.

In the present study, the greater alpha diversity in SC nodules was more pronounced in pH group B compared with groups A and C (Tables 2.4 and 2.5). In order to better cope with the low pH

environment, SC roots may actively and selectively recruit 'other' bacteria into nodules which may assist in pH-stress tolerance. The relative abundance of *Pseudomonas*, *Salmonella* and *Sinorhizobium* changed the most in SC nodules across the three pH groups (Figure 2.3). Previous research has shown that *Pseudomonas* contains many growth-promoting species that can play an important role in helping plants alleviate stresses such as salinity (Bano and Fatima 2009) or assist in conferring resistance to pathogens (de Zelicourt et al. 2013). In their review, Martinez-Hidalgo and Hirsch (2017) proposed that the rhizobia and 'other' (non-nodulating) bacteria act together within the root nodule to facilitate plant health and survival, particularly under stressful environmental conditions, the current work supports that theory. The work of Maxwell *et al.* (2012) also showed that SC plants are able to withstand acidic soils better than WC plants, since the total dry matter of *T. subterraneum* cv. Mt Barker was higher than that of *T. repens* cv. Nomad when grown in soils of pH 5.2–5.5. Changes to the nodule microbiome may contribute towards this observation.

The results showed that the nodule alpha diversity was different between the two host plants, with SC showing a strong trend in response to the pH of the external soil environment, but not WC. It is likely that the inherent variability within WC has decreased the statistical power for the WC response (Wakelin et al. 2018). This indicated that the clovers may be selectively filtering the nodule community in order to maximise the functional rewards from the nodule microbiome, this has been shown in other plants (Sørensen and Sessitsch 2007; Hardoim et al. 2008; Van Overbeek and Van Elsas 2008). Overall, the implications of these findings are: i) that the taxa in the nodule biome (other than rhizobia) may have functional significance and this should be explored and ii) that plants can modify the nodule biome in response to a stress such as low pH and this phenomenon should be explored for stress factors in addition to soil pH, such as, nutrient- or drought-stress. The question still remains as to whether this is cultivar or species dependent as only one cultivar of each host was explored. Further work on nodule microbiomes with more cultivars of SC and WC would determine whether the phenomenon is consistent across the species. Further work could also include a comparison of the root microbiome (endophytes), nodule microbiome and rhizosphere communities when the clovers are under stress to determine whether the nodule inhabitants are likely to be derived from endophytes or through direct recruitment from the rhizosphere.

The PCOs of the abundances of *Rhizobium* (Figure 2.4) and non-*Rhizobium* (Figure 2.5) genera across the samples showed a horseshoe effect. This horseshoe effect is likely caused by sampling a strong ecological gradient such as pH, but other environmental factors may also influence the bacterial/rhizobial populations. Horseshoe or arch effects commonly occur in data sets containing ecological gradients (i.e. spatial shifts in climate, soils and various environmental factors) due to the saturation property of distance metrics (Morton et al. 2017). This phenomenon is typified by a linear gradient that appears as a curve when projected onto a two-dimensional ordination space. Morton *et al.* (2017) performed a case study on the work of Lauber *et al.* (2009), which used

pyrosequencing to characterise bacterial populations from 88 soils on a continental scale. A classic horseshoe effect was observed in the correspondence analysis, showing a clear separation of the bacterial communities based on pH (Morton et al. 2017). If the nodule community was solely driven by the available taxa in the soil, then it would be expected for both hosts to have the same community structure – that they are different shows a host effect.

In this chapter, the effects of soil biogeography (spatial separation, soil classification and climate) on bacterial community were not evaluated, because the primary focus was soil pH and related physicochemical properties. Other soil properties such as C, N, P, K were also not empirically tested in this study. A study using *16S rRNA* amplicon sequencing (V3-V4 regions) of bacterial communities from 110 sites in New Zealand showed that soil bacterial communities differed more in response to changing specific soil properties than in response to changes in climate or increasing biogeographic distance (large spatial scales and land use types) (Hermans et al. 2017). Fierer and Jackson (2006) also emphasised the overwhelming effects of pH on the diversity and richness of bacterial communities in 98 soils across a continental scale. However, more recent research by Lammel *et al.* (2018) has identified that there are two distinct pH-related mechanisms driving prokaryotic community structures in agricultural soils: the direct effect and "spillover" (indirect) effects of pH. The authors sampled agricultural soils with a pH gradient of 4.0–6.0 in Brazil, amplified the V4 region of *16S rRNA* gene and found an increase in bacterial diversity with pH. The authors also stated that the indirect effects are highly relevant for some bacterial OTUs and consequently for the community structure; they proposed that indirect pH effects should be further investigated in microbial ecology (Lammel et al. 2018). Thus, further work on indirect pH effects, specifically on diversity and richness of *Rhizobium* spp., should be carried out.

2.4.3 *nodC* data

This chapter assessed the abundance and diversity of the *nodC* community in over 5,000 nodules of two closely-related clovers grown in soils with a broad pH range. This gene was chosen as a target for analysis because the *nodABC* genes are highly conserved in all rhizobia and their respective protein products are required for the synthesis of Nod factors that induce nodule formation (Barny et al. 1996). The *nodC* gene served as a proxy to study genotype variation as influenced by soil pH, but not necessarily as the link to a functional role in pH adaptation. NodC is an N-acetylglucosaminyl transferase (chitin synthase), which plays a functional role in making the chito-oligosaccharide chain (Spaink et al. 1994). This *nod* gene was chosen over other genes for the following reasons: i) work in New Zealand has shown that this gene was a single copy in most strains; ii) polymorphism within this gene was relatively high (Ridgway, unpublished data); and iii) polymorphisms could distinguish the main commercial inoculants used for subterranean (WSM1325) and white (TA1) clover. Thus, *nodC* diversity was likely to correlate well with the diversity of rhizobia in the system.

Overall, soil pH did not influence the number of *nodC* OTUs and reads in either clover species, although the abundance of some OTUs (representatives) were affected by soil pH and/or host species (Table 2.7). Some *nodC* genotypes were found to be pH specialists, others were host specialists, and few were found at higher frequencies in a particular host at a specific soil pH. If these genotypes are found to be related to strains isolated from nodules of SC or WC (which have demonstrated increased growth in acid/alkaline media; Chapter 3), there would be evidence for using those strains as inoculants for target clovers and/or soils of specific pH levels. In addition, the length of the NodC protein sequenced in the present study was much smaller (113 amino acids) than the complete NodC proteins (e.g. the WSM1325 NodC protein is 1,281 amino acids long). However, the portion of NodC protein sequenced contained key elements including one of three transmembrane domains, one of three cytoplasmic interface-leaning domains, the (E/D)DX-motif and the Q(R/Q)RW-motif from the *Sm*NodC protein alignment (Dorfmueller et al. 2014).

In this study only one of the several known *nod* genes was evaluated, albeit, for the first time to study the effects of soil pH on the diversity of *nodC* gene. However, to better understand nodulator diversity and abundance in clovers (or other legumes), other nod genes (*nodA*, *nodB* and/or *nodD*) could be included. For example, the diversity of *nodD* sequences amplified from a soil sample was compared with that of 134 isolates of *R. leguminosarum* bv. *trifolii* obtained from root nodules of white clover inoculated with the same soil (Zézé et al. 2001). The authors found that nodule isolates were of four *nodD* RFLP types, with 77% being of a single type. They also found the same four RFLP types among the clones from soil DNA, and stated that white clover selects specific genotypes from the available soil population (Zézé et al. 2001). Similarly, Young *et al.* (unpublished) have developed a system (QQAD – Quantitative Qualitative Amplicon Diversity) to assess the diversity of a *Rhizobium* nodule population. This involved amplification of *recA*, *rpoB*, *nodA* and *nodD* individually with *R. leguminosarum* bv. *trifolii* specific primers (Young, unpublished).

The commercial rhizobia strains TA1 and WSM1325 typified the two main groups of *nodC* OTUs (protein sequences) (Figure 2.6 and Table 2.7). In order to understand how different the two groups were, changes in amino acids between them were elucidated by *in silico* translation of the DNA sequence. The main difference between the protein sequences of TA1 and WSM1325 was an amino acid substitution of asparagine (polar side chain) to aspartic acid (electrically-charged side chain) (Δ30N$\rightarrow$D). Other non-conservative changes in amino acids between Groups 1 and 2 are arginine to glutamine (Δ44R$\rightarrow$Q), glycine to aspartic acid (Δ67G$\rightarrow$D) and threonine to alanine (Δ79T$\rightarrow$A). Some of the several conservative changes in amino acids between Groups 1 and 2 included leucine to methionine (Δ50L$\rightarrow$M), methionine to isoleucine (Δ60M$\rightarrow$I) and alanine to valine (Δ82A$\rightarrow$V), which are unlikely to be related to functional changes in the NodC protein.

The two groups distinguished the two most widely used commercial rhizobia inoculants in clover pastures in New Zealand, and the OTUs within each group are likely to include progeny of either

TA1 or WSM1325. There is also a possibility that they reflect horizontal transfer of the *nodC* gene from TA1 and/or WSM1325 to naturalised strains in pastoral soils. Studies have found high similarities of symbiotic (*nod* and *nif*) genes between the evolutionarily divergent α- and β-rhizobia, suggesting lateral transfer between strains (reviewed in Martinez-Hidalgo and Hirsch, 2017). However, although this was a comprehensive and recent review of research on the nodule microbiome, it did not well-represent *Trifolium* as a host. The review focused on studies that isolated nodule bacteria (by culturing), not studies that have used amplicon sequencing of nodule communities (Martínez-Hidalgo and Hirsch 2017). A more recent review by Andrews *et al.* (2018) also stressed the occurrence and importance of horizontal gene transfer between rhizobial genera. One of the studies reviewed therein found that the symbiotic genes (*nodC* and *nifH*) corresponding to *R. leguminosarum* bv. *trifolii* were shared across naturalised strains nodulating *T. repens* in alkaline soils in China (Zhang et al. 2016).

2.4.4 Soil properties and nodule microbiome

The observed correlations between soil pH and the other physicochemical properties of the soils used in this study were as expected because they have a high degree of collinearity. Soil pH is closely related to the solubility of some minerals[2], affecting their levels in soil and in turn their availability to plants (Table 2.8). Levels of calcium, magnesium and potassium (Ca^{2+}, Mg^{2+} and K^{+} are basic cations) increased with increasing soil pH. In contrast, Al^{3+}, an acid cation was found at higher levels in low pH soils, since Al complexes are soluble at low pH (Samac and Tesfaye 2003). Comprehensive studies on 40 forest soils (Blagodatskaya and Anderson 1998; Bååth and Anderson 2003), spanning a pH range of 3.0–6.0, demonstrated that soil pH and substrate (mineral) availability were important factors in determining soil microbial (bacterial and fungal) activities. However, in those studies the intensity of the response was also modulated by the plant with the effect of pH more pronounced under beech than under spruce. In both studies, variations in soil pH were the result of chemical characteristics of different types of soil (similar to the soils used in this chapter). Another study by Pietri and Brookes (2008) conducted on a 200 m natural, continuous pH gradient (pH 3.7–8.3) in a single soil type planted with barley in Rothamsted (UK) showed that microbial biomass and activity was affected by soil pH, with values stabilising at pH 5.0–7.0. The authors attributed this to more stable organic C, total N and Al concentrations within that pH range (Pietri and Brookes 2008).

Other factors (not measured in this research) such as land use change, fertiliser inputs or soil type may also have influenced microbial communities in the soils (Murty et al. 2002; Johnson et al. 2003; Lauber et al. 2008). However, these effects are likely to have had a more minor influence on soil bacterial communities than soil pH. For example, Lauber *et al.* (2008) used soils from 12 plots with four land-use types (hardwood and pine forests, cultivated and livestock pasture lands)

[2] Soil properties positively related to soil pH are: K, Ca and Mg. Those negatively related to soil pH are: Al, and organic matter. These will be collectively referred to as 'pH-related properties'.

and varying edaphic properties to assess bacterial (*16S rRNA* gene) and fungal (*18S rRNA* gene) community structure and distribution. The authors found that the phylogenetic distance between bacterial communities correlated with soil pH and soil texture, but not land-use type (Lauber et al. 2008). In a study with 32 agricultural soils planted to almond, cotton, grape, tomato and safflower, Johnson *et al.* (2003) found bacterial DNA fingerprints (*16S rRNA* gene and bacterial-specific *23S rRNA* gene) to be more similar to one another in samples collected within the same field than in samples collected from different fields (among the same crop species). The authors also reported that microbial community composition was strongly associated with crop type and soil properties (Johnson et al. 2003). However, the current study has focussed on the relationship between soil pH and the nodule biome. The effect of factors other than soil pH on the microbial communities in nodules was not examined but this would be an interesting area of further work, especially as research has suggested that plants can modify their interactions with microorganisms to assist in alleviating stresses such as salinity, heat, drought, and susceptibility to pathogens (Rodriguez et al. 2008; Nadeem et al. 2014).

2.4.5 The effect of soil properties on SDW and nodule numbers

Soil physicochemical properties influenced SDW and the number of nodules formed on the two clovers species. Fewer nodules were recovered from both clover species grown in soils with low (acidic) pH (Table 2.9 and Table 2.10). This was expected because previous work has also shown soil pH to have a marked effect on initiation of nodulation (Andrew 1978; Lin et al. 2012). There were differences in nodule numbers between pH groups A and C and also groups B and C for SC (Table 2.12), whereas there was only a trend between groups A and C ($p = 0.072$) for WC. This provided an added justification of splitting soils of pH 5.5 – 7.5 into two groups (B and C). The mechanisms for reduced nodule formation in low pH soils can be plant-driven (Andrew 1978), rhizobia-driven (Brockwell et al. 1991) or systemic, through effects on gene expression (Lin et al. 2012). Andrew (1978) stated that nodule formation was hindered in low pH soils due to lack of essential nutrients (Ca or P) or increased toxic Al^{3+} ions, as these negatively impact plant growth. In a study carried out across 84 sampling sites in New South Wales, Australia by Brockwell *et al.* (1991), the numbers of *Sinorhizobium meliloti* on six different *Medicago* species were found to be lower in soils of pH <6.0. The authors stated that the *Medicago* species were more tolerant of low pH than the rhizobia. Using qRT-PCR (quantitative reverse transcriptase real-time PCR), Lin *et al.* (2012) found that four nodule development genes were down-regulated during early nodule ontogenesis in pH 4.0 treated soybean plants compared with pH 6.0 treated plants, thus indicating a mechanism of systemic nodule inhibition.

Increased soil Ca^{2+} was related to more nodules in both SC and WC, which was supported by the work of Andrew (1978), which showed that Ca^{2+} improved nodulation in legumes especially at a soil pH range of 5.0–6.0. Results in the current study also showed that increased Al and OM were related to lower nodule numbers in both clover species. However, correlations of nodule

numbers with Ca, Al and OM are interrelated with soil pH and it is likely that soil pH is the major factor influencing nodule formation (see Footnote 2 – 'pH-related properties').

Higher soil P and K levels positively correlated with SDW in both clover species (Tables 2.9 and 2.10), because these minerals are essential to plant growth as previously noted in dairy, sheep and beef farming in New Zealand (Morton and Roberts 1999; Roberts and Morton 1999). Divito and Sadras (2014) performed a quantitative analysis on 63 studies related to nitrogen fixation, P, K and S with legumes including, but not restricted to, *T. repens*, *Glycine max*, *M. truncatula*, *Vicia faba* and *Phaseolus vulgaris.* The meta-analysis review (on pooled data across legume species, growing conditions and plant growth stage) found that deficiency of P, K and S reduced nodule growth and number to a greater extent than shoot mass (Divito and Sadras 2014). In addition, they found that when P, K and S were deficient, nodule activity decreased more than both shoot and nodule mass, which indicated a reduction in nodule productivity. The authors suggested that these effects were related to the direct effects of P, K and S on physiological and metabolic processes.

However, there are some caveats to consider when viewing the SDW results with measured soil physicochemical properties: i) only pairwise comparisons were made (which do not take into account the intricate relationships between individual soil properties); ii) the clovers were grown during winter for a period of only eight weeks; iii) the soil physicochemical properties were averaged across all soils; and iv) a statistically significant correlation does not necessarily relate to a biological causative effect.

2.5 Conclusions

- This work showed that there was a relationship between soil pH and the community of nodule bacteria in SC and WC. However, the alpha diversity of nodule bacteria was higher in SC than in WC. This work is the first to show that SC nodules recovered from highly acidic soils (pH group A) had higher *16S rRNA* alpha diversity than pH groups B or C.
- This work showed that although there was a relationship between pH and the nodule biome, clover species had a stronger effect on the beta diversity of nodule bacteria than any of the edaphic properties.
- This work also showed, for the first time, that there are significant differences in the relative abundance of nodule bacteria between SC and WC, with SC nodules harbouring relatively more non-*Rhizobium* bacteria than WC nodules, irrespective of soil pH.
- This work used *nodC* to explore genotypic diversity of rhizobia in the nodule and showed that the frequency of some *nodC* OTUs were higher in a particular clover species or pH range, which indicated pH adaptation/tolerance by naturalised rhizobia strains.

- New information on *nodC* diversity in New Zealand clovers was produced, with 353 partial *nodC* sequences identified from nodules of SC and WC. The data supported that *nodC* is a highly conserved gene and presents a likelihood that the naturalised strains are progeny of the commercial clover inoculants TA1 and WSM1325.
- As expected, soil pH affected nodule formation in both clover species – fewer nodules were formed when plants were grown in soils of low pH. This was also related to fewer nodules recovered from soils with high Al and OM, since pH, Al and OM are related soil properties.
- The soil properties that influenced SDW in both clovers were consistent (Ca, P and K) and strongly related to soil pH, i.e. dry weights were higher in soils with more Ca, P and K.

Finding acid- or alkali- tolerant *Rhizobium* strains could help in targeted delivery of those strains to soils that are acidic or alkaline, without the necessity of adding lime or other expensive soil amendments. This will be investigated in Chapter 3.

3 Investigating pH-adaptation in naturalised strains of clover rhizobia

3.1 Introduction

pH is a first-order factor influencing the niche preferences of soil microorganisms and has been convincingly shown to be a key driver of soil bacterial communities (Fierer and Jackson 2006; Lauber et al. 2009; Rousk et al. 2010), but it also affects chemical speciation of soil elements such as Ca, Mg, Na, Al, Cu, Zn, Mn and hence the indirect effects of pH (Lammel et al. 2018) on host plant and/or bacterial communities (structure, physiology, functions). In Chapter 2, new evidence showed that the bacterial community in the nodule was influenced by soil pH, with high *nodC* diversity demonstrated for the rhizobia community. Changes to diversity of nodule rhizobia may reflect genetic changes in the free-living rhizobia population from which they are recruited. There was some evidence for specific *nodC* genotypes that were preferentially associated with a particular host and/or soil pH, and this may also indicate adaptation of free-living rhizobia to that soil pH.

There are reported examples of rhizobia adaptation to soil pH extremes (Brockwell et al. 1991; Denton et al. 2002; Watkin et al. 2003). Watkin *et al.* (2003) showed using laboratory screening that the growth of *Rhizobium leguminosarum* strain (WSM409) was higher in media of pH 4.8 compared with growth of TA1 strain. They termed these strains acid-tolerant (WSM409) and acid-sensitive (TA1). The results of Denton *et al.* (2002) indicated that naturalised rhizobia strains (isolated from soils of pH 8.0 and 8.5) were more tolerant to alkaline soil conditions and were better able to nodulate clovers (*Trifolium alexandrinum*, *T. purpureum* and *T. resupinatum*) compared with commercial strains (TA1 and WU95). The authors reported that commercial strains of rhizobia were poorly competitive under alkaline conditions, even compared with small, naturalised populations of *R. leguminosarum* bv. *trifolii* (Denton et al. 2002). Experiments on strains of *Sinorhizobium meliloti* with *M. sativa* growing in solution culture at pH_{CaCl2} 4.8 or 5.0 have shown that some strains are better at nodulating lucerne plants at acidic pH than others, and the authors suggested that large gains to plant yield could potentially be made by selecting and replacing the commercial *Rhizobium* strains (Charman et al. 2008; Humphries et al. 2009). Thus, naturalised strains that are pH adapted may have the potential to better compete and/or persist in acidic or alkaline soils compared with introduced commercial strains.

For rhizobia, acidic soils are a problem in productive settings, negatively affecting clover-rhizobia symbiosis and thereby clover growth (Watkin et al. 2000; Dilworth et al. 2001). In New Zealand, this problem is most significant in high country farming, and often can only be addressed through extensive liming (Lowther and Kerr 2011). This can be costly for farmers, particularly those on steep or hill country farms where lime can only readily be applied by fixed-wing aircraft or helicopters which incur considerable expense to farmers. In an Australian field study, addition of

lime increased soil pH, reduced extractable aluminium concentrations below phytotoxic concentrations and increased nodulation of subterranean clover (Unkovich et al. 1996). Although it is possible to manage acid soils using liming, it is difficult to implement in the hill country pastoral farms and an alternative approach such as using acid-adapted rhizobia strains as clover inoculants could prove to be more feasible. Hence, there is an opportunity to improve the growth of clover in high country soil by using strains of rhizobia that have tolerance to soil pH extremes. This need provides the applied impetus for research to understand the drivers of soil pH adaptation in rhizobia.

The goal of this chapter was to determine whether there was evidence of pH adaptation in SC and WC nodulating strains of rhizobia and whether this was related to the pH of the soil of origin. According to the 'intermediate disturbance hypothesis', which was first described by Connell (1978) "diversity is higher when disturbances are intermediate on the scales of frequency and intensity". Since its inception, the hypothesis has been applied to numerous studies on plant, animal and microbial communities. Connell's hypothesis discussed diversity, disturbance and intensity; these can be related to population of pH-adapted strains, pH-stress and acidity or alkalinity. Thus, the predictions from this were: i) strains isolated from soils in pH groups A or C (section 2.2.2) will be adapted to grow at low and high pH, respectively; and ii) strains isolated from soils in pH group B will be able to grow in a broad pH range.

For the purpose of this research, 'pH-adaptation' by rhizobia strains was defined as their relative ability to grow optimally in liquid media at either low pH (< 5.5), high pH (> 7.5) or across a broad pH range (4.5–9.0). To achieve this goal the following objectives were developed: 1) To recover a representative collection of rhizobia from a sub-set of 12 soils spanning the breadth of the selected pH range, 2) To develop a robust bioassay to investigate the ability of rhizobia strains to grow in liquid culture at a range of pH and 3) To use the bioassay to identify rhizobia strains adapted to low (<pH 5.5), high (pH >7.5) or a broad range of pH.

3.2 Materials and methods

3.2.1 Soil baiting for recovery of rhizobia

A selection of 12 soils (Table 3.1) was made from all sampled soils (section 2.2.1) and used for a soil baiting experiment from which bacteria colonising the nodules were recovered into culture. The criteria for selecting these 12 soils were: spanning a pH range from 4.9–7.5; geographically diverse sampling locations; representing different livestock types (deer, dairy, sheep and beef) within pastoral grazing farms.

Table 3.1: Details of soils selected for a rhizobia bating experiment, showing the sampling location, land use and soil pH results after sampling. Full soil test results in Appendix 3: Supplementary data.

Paddock name	Location	pH	pH group
G13+BH29	Westport - Cape Foulwind	4.9	Group A (pH 4.3 – 5.3)
WE_C22	Dobson - Weka	5.2	
Bills	Wairoa - Panekiri	5.2	
AR_120	Taupo - Aratiatia	5.5	
TO_39	Westport - Totara	5.8	Group B (pH 5.5 – 6.1)
DG_D16	Manapouri - Duncraigen	6.0	
TO_C06	Westport - Totara	6.1	
CH_F38	Westport - Cape Foulwind	6.1	
SU_37	Dobson - Souters	6.5	Group C (pH 6.5 – 7.5)
ER_38	Reporoa - Endevour	6.6	
SiberiaWatch	Napier - Ahuriri	7.1	
Pidgeon13	Napier - Ahuriri	7.5	

Each soil was sieved and homogenised prior to taking a sub-sample. For each of the sites, three 100 g sub-samples of soil were mixed with 20 g coarse pumice (sterilised by autoclaving) and placed into each of three replicate 200 mL pottles. For free drainage to occur, five small holes were made at the bottom of each pottle by piercing with a hot needle. Seeds were surface-sterilised, sown and thinned as per section 2.2.2. Rhizobia were recovered from the nodules of clover plants grown in pottles containing each of the 12 soils.

Each clover species was planted and rhizobia recovered from nodules. SC was sown on 30 December 2015 and WC on 27 January 2016. Pottles were placed in large trays in a randomised complete block design in the Lincoln University shadehouse and watered lightly when needed. The clovers were harvested after 6 weeks growth; i.e. on 15 February 2016 (SC) and 14 March 2016 (WC). The average maximum recorded temperatures in the shadehouse for January, February and March were 22.4°C, 28.5°C and 21.2°C, respectively. The average minimum recorded temperatures in the shadehouse for January, February and March were 11.4°C, 12.8°C

and 9.8°C, respectively. At harvest, the number of plants in each pottle was recorded (max = 3) and the roots were washed in tap water to remove soil and stored in plastic containers at 7°C until all samples were ready to be sterilised (maximum of 3 d).

3.2.2 Isolation of rhizobia from nodules

All nodules were surface sterilised as described in section 2.2.2. To generate a culture collection, the rhizobia from two nodules per plant (randomly chosen based on representatives of size, colour and position on root) from each pottle were recovered by crushing each nodule separately with a sterile glass rod and streaking the bacteria onto separate yeast mannitol agar (YMA, Appendix 1: Media and reagents) plates. The YMA plates were sealed with plastic film and incubated in the dark at 25°C for 3–4 days. After incubation, single colonies of each rhizobia strain were sub-cultured onto fresh YMA plates a further two times to ensure purity. From the final subculture, a single colony was picked using a 10 µL sterile disposable loop (Biologix Group Ltd.) and inoculated into 1 mL of yeast mannitol broth (YMB, Appendix 1: Media and reagents). The tubes were incubated at 28°C for 24 h at 200 rpm in a shaking incubator (Labnet Problot125 Hybridization Oven, Labnet International, Inc.), after which two 500 µL aliquots from each tube were mixed with 37 µL dimethyl sulfoxide (DMSO) and stored at -80°C.

Strain labelling was done using the following convention: Soil/Replicate ID–Plant ID–Nodule ID, e.g. strain 3P3N2 refers to replicate 1 from soil WE_C22, the 3rd plant and the 2nd nodule.

3.2.3 Most probable number (MPN) quantification of rhizobia in each soil

The standard operating protocol for evaluating MPNs was modified from (Brockwell 1963; Marshall et al. 1993) by Hayley Ridgway, Céline Blond and Dalin Brown (unpublished). Surface-sterilised (section 2.2.2), air-dried seeds of each clover species were sown in approximately 30 g or 4.5–5.5 mL of firmly packed sterile vermiculite in separate plastic tubes (one seed per tube). Due to the variable size of the two clover species, 50 mL tubes were used for SC and 15 mL tubes were used for WC. Each tube contained 2.5 mL of N-free McKnight's nutrient solution (Appendix 1: Media and reagents). The seeds were germinated for 5 d in a plant growth chamber (Adaptis, Controlled Environments Ltd, Canada) with a 16 h photoperiod (400 µmol photons $m^{-2}s^{-1}$) at a constant 22°C.

To prepare the soil suspensions, each of the 12 soils were diluted by placing 10 g of soil in 90 mL of 0.85% w/v saline solution. To homogenise the suspension, the bottles were placed on a flask shaker (Stuart, Biddy Scientific Limited, Staffordshire, UK) for 15 min at maximum speed. The first soil suspension (10^0) was serially diluted (10-fold) to 10^{-5} with 0.85% w/v saline. The base of each germinated seedling was inoculated with 1 mL (SC) or 100 µL (WC) of a soil suspension, with five replicates per dilution. Cellophane was wrapped over the tube opening and a small hole was

made for gas transfer. The plants were incubated at 22°C for 42 d in a plant growth chamber (Adaptis, Controlled Environments Ltd, Canada) with a 16 h photoperiod (400 µmol photons m^{-2} s^{-1}), after which their roots were assessed for presence or absence of nodules. The average number of nodulated plants for each dilution was entered into the MPN calculator and MPN was determined (Hurley and Roscoe 1983). Correlations of the MPN counts with soil physicochemical properties and SDW were done in R (v 3.4.3).

3.2.4 Taxonomic identification of strains by *16S rRNA* gene sequencing

Approximately 10 µL of culture (from the stored stock, Section 3.2.2) was collected in a sterile loop and streaked onto a fresh YMA plate to obtain single colonies, after an incubation period of 3–4 days. Using a sterile pipette tip, a small amount of a single colony was picked and mixed into the PCR master mix (10 µL of DreamTaq PCR mix, 1 µL each of forward and reverse primers (5 µM) (Appendix 2: Supplementary sequencing data) and 7 µL of H_2O). Primers were designed as described by Tan *et al.* (2012).

The PCR protocol was the same as described in section 2.2.3. The PCR products were separated by electrophoresis on a 1% agarose gel in 1× TAE buffer at 100 V for 45 min. Bands of the appropriate size and concentration were directly sequenced using the reverse primer (R1494) at the Lincoln University Sequencing Facility. The sequences obtained were viewed and trimmed using FinchTV v 1.4.0 (Geospiza, Inc.) to remove ambiguous sequence ends. The phylogenetic identity was determined by comparing sequences to those of known origin using the unrestricted nucleotide basic local alignment search tool (BLASTn) against deposited sequences in the NCBI GenBank database (www.ncbi.nlm.nih.gov/genbank/). Sequences which had less than 98% sequence identity to the genus *Rhizobium* were re-amplified and re-sequenced.

3.2.5 Development and optimisation of a liquid culture pH assay

Three pilot studies were performed to develop the methodology and parameters for the main bioassay experiment.

3.2.5.1 Pilot study 1 – assessing strain growth on pH-adjusted broth

For the first pilot study, five strains of WC rhizobia were chosen. These were two strains from pH group A soils (1P1N1 and 2P1N1), one from group B soil (6P1N1) and two from group C soils (47P1N1 and 48P1N1). A primary culture was produced by placing 20 µL of each strain (from the stored culture at -80°C) in 800 µL of sterile YMB, in duplicate and the tubes shaken at 200 rpm (brand of shaker) at 27°C for 48 h. To ensure all cultures had reached a similar growth stage after 48 h (OD of 0.8–1.0) the OD_{600nm} was measured using a Jenway 6305 spectrophotometer. YMB was prepared and adjusted to the following pH values by adding either 5M HCl or 5M NaOH: 4.5, 5.0, 5.5, 6.5, 7.5 and 8.0 (measured using the Infetcom, Japan S2K712 pH meter). An 800 µL

aliquot of each pH-adjusted YMB was added to a 1.7 mL tube. For each pH value, 16 tubes were prepared (5 strains × 3 replicates + 1 control). For each strain, 20 µL of the 48 h primary culture was added to each of the 1.7 mL tubes containing the pH-adjusted YMB (18 tubes/strain: 6 pH levels × 3 replicates). The tubes were randomised in racks and shaken at 200 rpm (Labnet shaking incubator) at 27°C for 48 h. After this time, 200 µL from each tube was aliquoted into an individual well of a 96-well Jet Biofil® (Jet Bio-Filtration Co. Ltd., China) tissue culture plate and the OD_{600nm} was measured using the Thermo Scientific Multiskan® GO plate-reader and the measurements were recorded using the SkanIt Software version 3.2 (Thermo Scientific).

3.2.5.2 Pilot study 2 – optimising assay incubation time

This pilot study was similar to the first, except that for each strain the OD_{600nm} was measured at three time points (24 h, 30 h and 48 h) and a larger number of pH levels were included. Only three of the five strains were used for this study (1P1N1, 6P1N1 and 47P1N1). Seven pH values were used, and these were 4.6, 5.1, 6.3, 7.0, 8.5, 9.3 and 10.5. There were four replicates/strain/pH. The primary culture strains were allowed to grow for 30 h in the shaking incubator with the same settings as pilot study 1. The OD_{600nm} was measured for each strain at the end of this period and the procedure described previously was followed to inoculate the pH amended YMB media. OD_{600nm} readings were taken after 24, 30 and 48 h incubation. The time points were used to estimate the rate of growth by calculating the Δ OD_{600nm} (OD_{600nm} 48 h - OD_{600nm} 30 h; OD_{600nm} 30 h - OD_{600nm} 24h; OD_{600nm} 48 h - OD_{600nm} 24h).

3.2.5.3 Pilot study 3 – determination of broth pH at endpoint

An additional test was conducted to assess whether some strains altered the pH of the broths (two of the low pH and two of the high pH broths) at the end of the growth period. Two randomly chosen strains each from SC and WC were used, as well as the commercial strains TA1 and WSM1325. A 20 µL aliquot of the strains was inoculated into 800 µL YMB each of pH 4.5, 5.4, 8.0 and 9.1 (in duplicates) and shaken at 200 rpm in an incubator at 28°C for 40 h. The pH of the broths was measured at the end of the 40 h growth period.

3.2.5.4 Main pH assay

For the final optimised assay, the pH values of the media were (+/- 0.1): 4.5, 5.4, 5.8, 6.5, 7.5, 8.0 and 9.0. For each strain, 200 µL (-80°C culture) was added to 4 mL of sterile YMB in a 15 mL tube and shaken at 200 rpm in an incubator at 28°C for 26 h. Of this culture, 1 mL was used to measure the initial OD_{600nm}, thus ensuring that all strains had reached the same growth stage (OD 0.8–1.0). A 20 µL aliquot of this culture was added to 800 µL pH-amended YMB in separate wells of 1.2 mL/well volume 96-well plates (Masterblock®, Greiner Bio-One). Each plate was sealed with Adhesive PCR sealing foil sheets (Thermo Scientific), which were pierced at each time point to take OD readings and this allowed gas exchange to occur. A maximum of 36 strains were assessed

at the same time (a batch) with three replicates/strain/pH, resulting in a total of nine 96-well plates (each with 12 strains, seven pH levels and negative controls). The strains and pH amended YMB were completely randomised in each 96-well plate. The OD_{600nm} measurements were taken at 24, 38 and 48 h after pH-amended YMB was inoculated with the strains, by taking 200 µL of culture per well and transferring into a Jet Biofil® tissue culture plate. Measurements were corrected by subtracting the control values (OD_{600nm} of pH-amended YMB without inoculation) to produce corrected OD values.

To complete the assessment of all 138 WC strains, the experiment was split into five batches, with 10 strains used as internal standards between batches. A final (6^{th}) batch comprised those strains exhibiting divergent growth patterns for any pH level, to confirm their growth patterns. The main assay was repeated for 161 SC strains with two commercial strains (TA1 and WSM1325) and three WC strains included as controls. The SC assessment was split into seven batches with six strains used as internal standards between batches. A final, 8^{th} batch comprised those strains exhibiting divergent growth patterns for the pH range, to confirm their growth patterns.

3.2.5.5 Statistical analysis

A t-test was performed (R v3.4.3) on the 48 h OD measurements of the 10 strains used as internal controls across the six batches of WC strains to determine whether there were any experimental differences between them (Chapter 3 supplementary results). T-tests are used to determine the probability of difference between normally distributed populations with unknown variances and with small sample sizes. Between pairs of experiments, there are overlapping strains, but these change over the complete set of experiments. The strains were also randomised across the batches to avoid bias of all strains from the same soil growing together in one batch.

Another t-test was carried out to determine how different the 6^{th} batch was from the mean of batches 1–5 (Chapter 3 supplementary results). The means were adjusted in the same way as for the previous t-test. The result further confirms what was observed from the previous t-test, wherein batch 6 had a slightly higher mean, but not significantly different from the combined means of batches 1–5. A t-test was performed on the OD measurements (across the three time points) of the six internal control strains across seven batches of SC strains to determine whether there were any experimental differences between them. The internal control strains were not present in batch 1 and thus data from that batch could not be used for the t-test. The t-test results for batches of SC strains show that growth of strains from some batches was significantly different from other batches (Chapter 3 supplementary results), but because strains were randomised across the batches, general conclusions can still be drawn from the data without incurring biases.

The media optimum pH for each strain (WC and SC), at each time point was determined by observing which pH-amended broth caused the highest growth (maximum OD_{600nm} reading). For WC strains, the media optimum pH values at 38 h and 48 h were significantly correlated ($r = 0.4321$, $P < 0.0001$), and thus averaged to obtain a single optimum pH value for each strain. For the SC strains, the media optimum pH values at 40 h and 50 h were significantly correlated ($r = 0.5611$, $P < 0.0001$), and were also averaged for each strain. Growth data for SC and WC strains was visualised using 'ggplot2' package (Wickham 2011) in R, random-intercepts models were used to rank growth of strains (R v3.4.3) and Pearson's correlation analyses were performed between media optimum pH and soil physicochemical factors (R v3.4.3).

3.3 Results

3.3.1 Properties of selected soils

Soil physicochemical properties are not independent of one another, and there are known relationships between pH and mineral content, soil type, etc. The Pearson's correlations of pH with other properties of the 12 soils used for baiting are shown in Table 3.2. Soil properties in blue have significant positive correlations, whereas those in red have significant negative correlations with pH.

Table 3.2: Pearson's correlations of soil pH with other measured soil physicochemical parameters, and their corresponding p values. Values with * are significant at 95%.

	Soil pH	
	Pearson's r	p value
Olsen Phosphorus	0.5298	0.0009 *
Potassium	0.4658	0.0042 *
Calcium	0.8977	< 0.0001 *
Magnesium	0.6353	< 0.0001 *
Aluminium	-0.7953	< 0.0001 *
Organic Matter	-0.5727	0.0003 *

3.3.2 The abundance of free living rhizobia in the soil (MPN) and correlation with measured soil properties

For each soil, the MPN counts (most probable number/g of soil) for each clover species is shown in Table 3.3. For both clover species, the MPN counts were significantly correlated with soil pH ($r = 0.5659$, $p < 0.0001$) (Table 3.4). MPNs also had significant positive correlations with Olsen P ($r = 0.3517$, $p = 0.0024$), K ($r = 0.4098$, $p = 0.0003$), Ca ($r = 0.3453$, $p = 0.0029$) and Mg ($r = 0.4145$, $p = 0.0002$). Statistically significant negative correlations were present with Al ($r = -0.5536$, $p < 0.0001$), and OM ($r = -0.2406$, $p = 0.0417$). Figure 3.1 is a visual representation of the data for the relationship between MPN and soil pH, which shows the positive correlation of MPN with soil pH.

Table 3.3: Most Probable Number counts (CFU/g) for the two clover species cultured in 12 selected NZ soils of contrasting pH. SC = Subterranean clover, WC = White clover.

		MPN (CFU/g of soil)	
Soil	**Soil pH**	**SC**	**WC**
G13+BH29	4.9	5.50×10^2	1.20×10^3
Bills	5.2	3.30×10^3	3.30×10^3
WE_C22	5.2	4.00×10^2	1.20×10^3
AR_120	5.5	6.10×10^3	3.30×10^4
TO_39	5.8	8.00×10^3	5.80×10^3
DG_D16	6.0	3.70×10^5	2.80×10^4
TO_C6	6.1	4.90×10^5	2.30×10^4
CH_F38	6.1	1.20×10^4	2.10×10^5
SU_37	6.5	5.20×10^5	7.10×10^3
ER_38	6.6	2.50×10^5	5.90×10^4
SiberiaWatch	7.1	2.10×10^5	3.10×10^5
Pidgeon13	7.5	2.10×10^3	2.70×10^4

Table 3.4: Pearson's correlations of $\log_{10}$ MPN counts of both clover species with measured soil physicochemical parameters, and their corresponding p values. Values with * are significant at 95%. Properties in blue are positively correlated with MPN and those in red are negatively correlated with MPN.

	$\log_{10}$ MPN (cells/g of soil)	
	Pearson's r	p value
Soil pH	0.5659	< 0.0001 *
Olsen Phosphorus	0.3517	0.0024 *
Potassium	0.4098	0.0003 *
Calcium	0.3453	0.0029 *
Magnesium	0.4145	0.0002 *
Aluminium	-0.5536	< 0.0001 *
Organic Matter	-0.2406	0.0417 *

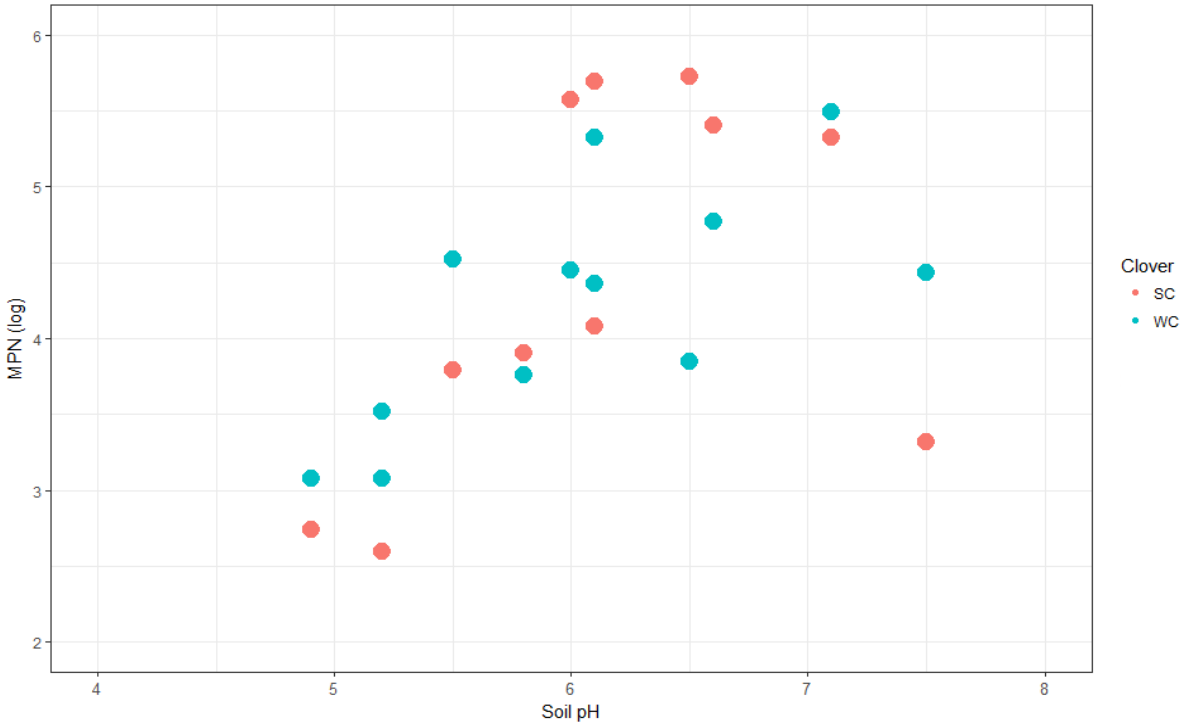

Figure 3.1: MPN counts (log scale) plotted against pH of the 12 soils for both SC (red) and WC (teal). SC = Subterranean Clover, WC = White Clover.

3.3.3 Identity of the bacteria recovered from the 12 soils

A total of 223 bacteria were recovered from the nodules of SC and 145 from the nodules of WC. PCR products of appropriate size for the *16S rRNA* gene portion were amplified (~700 bp) from all strains. The BLAST results from comparison of the DNA sequence of the *16S rRNA* from the recovered strains with those on GenBank showed that 161 SC strains (72%) and 138 WC strains (95%) had a sequence similarity of > 98% to *Rhizobium* spp.. This confirmed a collection of 161 and 138 strains of *Rhizobium* spp. from SC and WC, respectively. The number of strains of each clover recovered from soils in pH groups A, B and C are shown in Table 3.5.

Table 3.5: Number of strains of *Rhizobium* spp. cultured from nodules of plants grown in soils from pH groups A, B and C.

	Subterranean Clover (SC)	**White Clover (WC)**
Group A (pH 4.3 – 5.3)	61	29
Group B (pH 5.5 – 6.1)	52	57
Group C (pH 6.5 – 7.5)	48	52

A small number of recovered bacteria were not confirmed as *Rhizobium* spp.. From SC nodules, the 'other' genera found were: *Bacillus* (n = 1), *Herbaspirillum* (n = 1), *Ochrobactrum* (n = 2), *Staphylococcus* (n = 2) and Undefined (n = 56) (Figure 3.2). 'Other' genera identified from WC nodules were: *Neorhizobium* (n = 1), *Ochrobactrum* (n = 2), *Sinorhizobium* (n = 1) and Undefined (n = 3) (Figure 3.3). Undefined strains were either of low quality or ambiguous sequences.

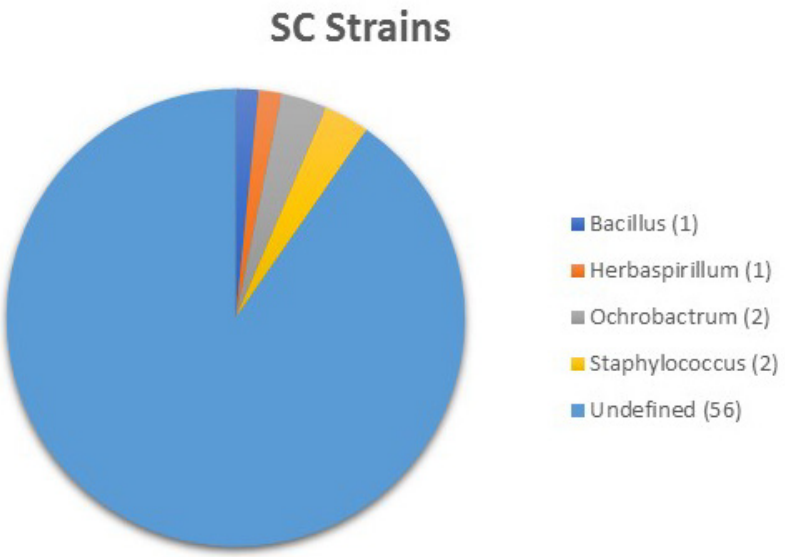

Figure 3.2: Identity of 'other' genera cultured from nodules of subterranean clover (SC). Number of strains in parentheses.

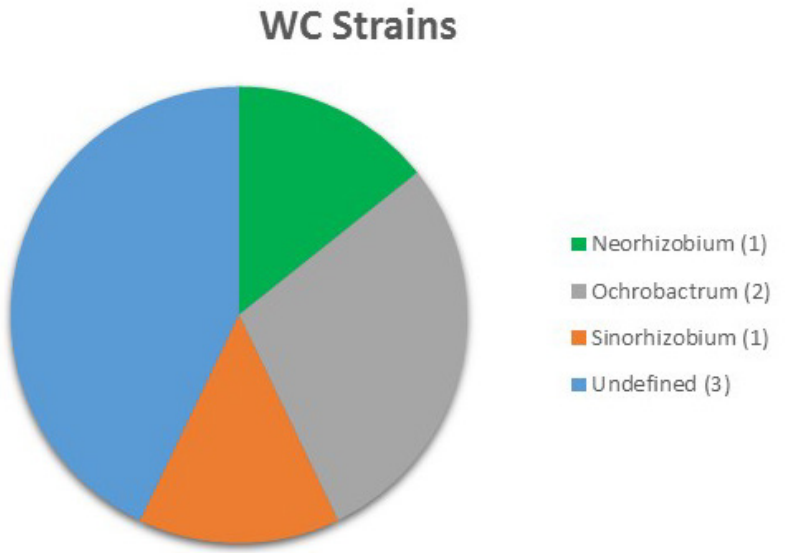

Figure 3.3: Identity of 'other' genera cultured from nodules of white clover (WC). Number of strains in parentheses.

3.3.4 Identification of rhizobia with tolerance to high, low or a broad range of pH

Pilot studies 1 and 2 – The results from these pilot studies (Chapter 3 supplementary results Figures A2 and A3) identified the broth pH values for the main assay, which meant excluding broths above pH 9. They further confirmed that at the time points for measuring OD, the strains were in the growth (log) phase at the end of 48 h, and had not entered the stationary phase. The optimised broth pH and duration was used to examine the effect of strain growth on broth pH in pilot study 3.

Pilot study 3 – Growth of bacteria in liquid culture can modify the starting pH through the secretion of metabolites. The results showed that the change in pH of the broth cultures, after 40h of growth, was consistent among the six strains used for testing, which indicated that there was little strain-to-strain variation. The pH of acidic broths had minimal or no change after 40 h of continuous growth. The pH of the alkaline broths was reduced (acidified) by the strains (Table 3.6). As acidification of the media may be one of the mechanisms by which the rhizobia have adapted to high soil pH, the decision was made not to add buffers to the liquid growth medium.

Table 3.6: Actual pH of the liquid culture after 40 h growth. Values are averaged for the duplicate samples for each strain.

	pH of liquid culture after 40 h growth					
	Subterranean clover			White clover		
pH of broth at start	**WSM1325**	**6P1N1**	**40P1N1**	**TA1**	**1P1N1**	**43P2N2**
4.5	4.7	4.7	4.7	4.7	4.6	4.7
5.4	5.6	5.3	5.4	5.7	5.6	5.8
8.0	6.7	6.6	6.6	6.8	6.6	6.7
9.1	7.2	7.0	7.0	7.1	7.0	7.0

Main assay – For each batch (maximum 36 strains) of SC and WC strains, a graph was plotted for the growth over the three time points in each pH-amended YMB medium. This allowed the identification of strains exhibiting growth patterns divergent from the mean. An example of this data is shown in Figure 3.4. Data from all three replicates was plotted to show variation of growth between the replicates for each strain. The final OD was approximately half of the 'mother culture' (section 3.2.5.4) because the concentration of bacteria added to each well was 50% of the original. This was done to ensure there was no depletion of the substrate over 48 h.

Figure 3.4: Data from one batch of white clover strains cultured in the seven pH-amended broths. The OD_{600nm} readings are shown for each of the three time points. Coloured dots represent the pH groups of soils of origin • Group A (pH 4.3 – 5.3), • Group B (pH 5.5 – 6.1), • Group C (pH 6.5 – 7.5).

All 138 WC strains were ranked according to their deviation from the mean, for the 48h OD measurements. The plots in Figure 3.5 show fitted linear models with random effects for the strains growing at either low (4.5, 5.4 and 5.8), medium (6.5 and 7.5) or high pH (8.0 and 9.0), where strains above the central horizontal line display significantly different growth than those below the horizontal line (vertical lines are standard error bars). The assumption is that the strain effects were normally distributed. Since the experiment had not been independently replicated multiple times, there was no statistical difference between consecutive strains. However, the inclusion of internal standards meant that strains to the far right were a best estimate (Alasdair Noble pers. comm, 2016) of those that will consistently grow better than the majority at a specific pH. The pH assay as presented comprised over 7,000 individual data points and without repeating the experiment multiple times, it was not possible to estimate individual means and, thus, significant differences. However, the growth assay was replicated for strains that had shown growth divergent from the majority of the strains and that had been selected for further assays. The replicated assay on 18 WC and 17 SC strains included the internal controls, two 'standard strains' and the two commercial strains.

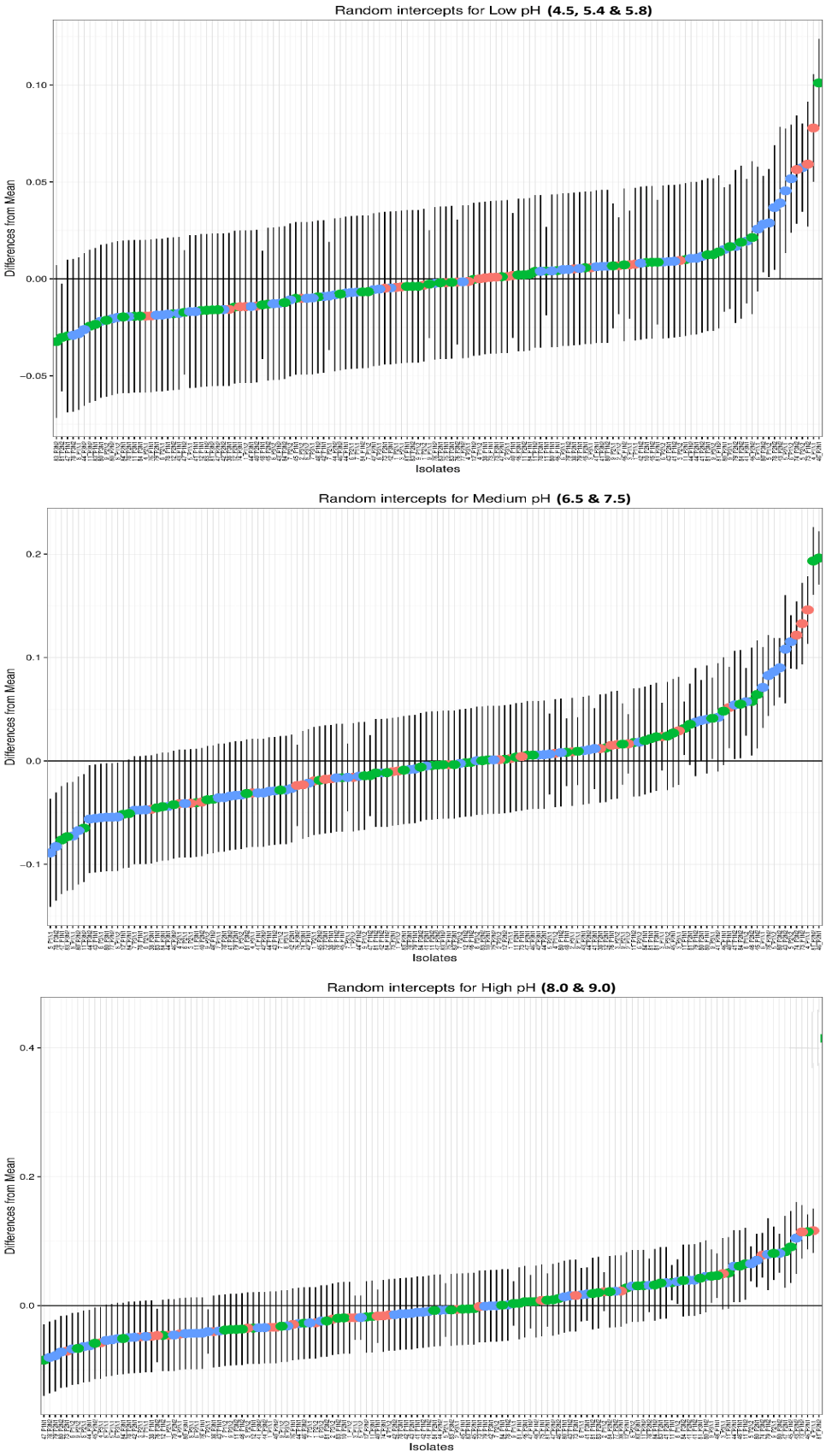

Figure 3.5: Fitted linear models for WC strains, with three groupings for media pH (low, medium and high). Strains on the x-axis and difference from the mean on the y-axis. Coloured dots represent the pH groups of soils of origin: • Group A (pH 4.3 – 5.3), • Group B (pH 5.5 – 6.1), • Group C (pH 6.5 – 7.5).

3.3.4.1 Relationship between media optimum pH and soil pH

The media optimum pH values for strains of SC and WC were positively correlated with the pH of the soil from which they were isolated (Figure 3.6). At the 95% CI, only SC strains were significant ($r = 0.667$, $p = 0.018$) but WC strains showed a strong trend ($r = 0.567$, $p = 0.054$). There was no correlation between media optimum pH and soil pH when data from strains originating from soils of pH 7.1 and 7.5 were removed (SC: $r = 0.107$, $p = 0.768$; WC: $r = 0.092$, $p = 0.800$). Other soil physicochemical properties also had significant correlations with the media optimum pH values for SC and WC strains. These correlations are shown in Chapter 3 supplementary results.

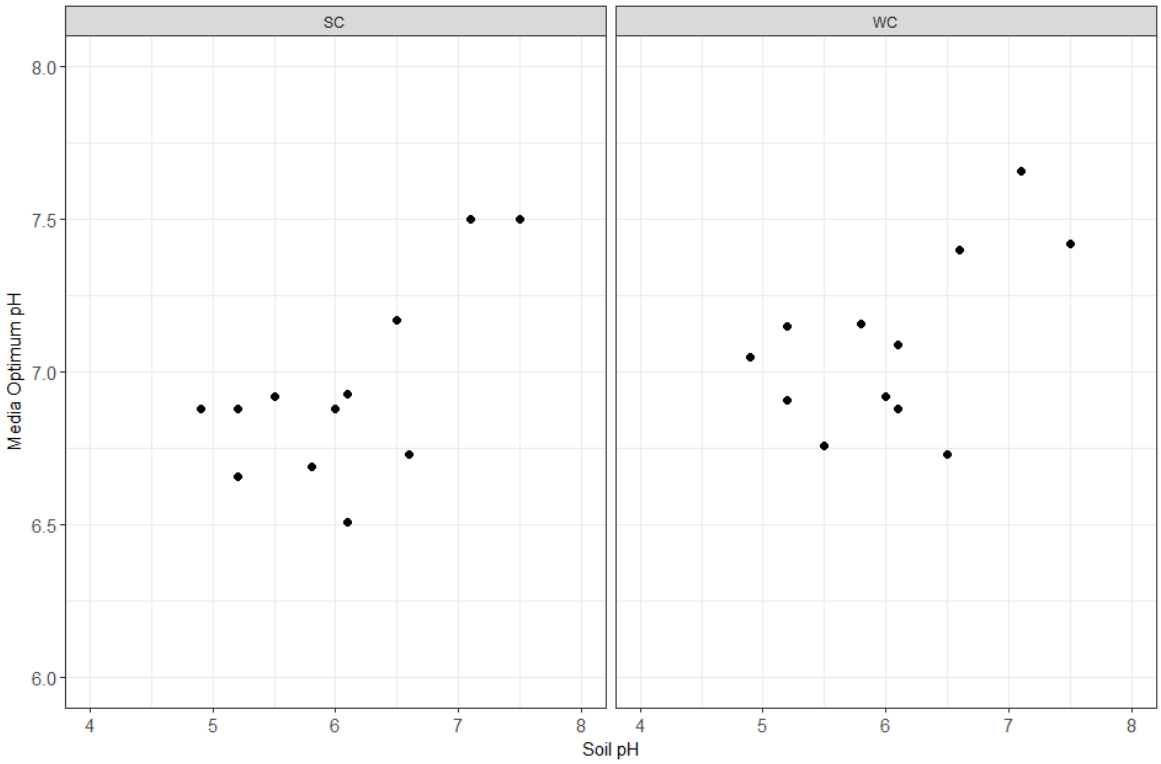

Figure 3.6: Correlation between the optimum pH in media and the soil pH of origin for strains isolated from the two clover species. SC = Subterranean Clover, WC = White Clover.

3.3.4.2 Selection of strains with greatest growth at specific pH

Strain selection was made on the basis of the relationship between the growth data observations from all batches to the random effects models (ranking). Strains with faster than average growth rates from the WC (n = 18, Table 3.7) and SC (n = 17, Table 3.8), along with strains of average growth rate (n = 2) and commercial strains (n = 2) were selected for further work. The 'average' strains were chosen because their difference from the mean was very close to zero in the fitted linear models (Figure 3.5).

From the 18 WC strains (13% of screened strains) displaying faster-growth, four strains were from soil pH group A, 10 from group B and four from group C. With the exception of three strains (1P1N1 – acid adapted, 47P2N1 and 81P2N2 – alkaline adapted) there was no relationship

between ‘pH-adaptation’ in WC strains and the pH of the soil of origin (Table 3.7). Strains 74P3N2, 4P1N1, 46P2N1 and 73P1N2 showed faster than average growth across a broad pH range. Three strains (5P2N2, 43P2N2 and 80P3N2) which originated from group B soils demonstrated adaptation to a broad pH range. The selected WC strains (and commercial strains) were further characterised based on their utilisation of different carbon sources (Chapter 4).

Table 3.7: Rhizobia strains isolated from White Clover (WC) with faster than average growth rate in pH adjusted liquid culture, showing media pH where strains exhibited above average growth rate (presence of **+**) and the pH of soil of origin (red = Group A, green = Group B, blue = Group C). Strain 9P1N1 was included because it showed average growth across the pH range.

	pH where strains demonstrate high growth							
WC Strains	**4.5**	**5.4**	**5.8**	**6.5**	**7.5**	**8.0**	**9.0**	**Soil of origin pH**
1P1N1	+							4.9
41P3N2	+							5.8
79P2N2	+							6.0
84P2N2		+						7.5
6P1N2	+	+	+	+				6.1
5P1N2	+	+		+				5.8
5P2N1			+					5.8
41P1N2			+					5.8
74P3N2	+	+	+	+	+		+	5.2
78P2N2			+	+				6.1
5P2N2		+	+	+	+	+		5.8
43P2N2		+	+	+	+	+		6.0
4P1N1		+	+	+	+	+	+	5.5
46P2N1		+	+	+	+	+	+	6.5
73P1N2		+	+	+	+	+	+	4.9
80P3N2			+	+		+	+	6.1
81P2N2					+	+	+	6.6
47P2N1							+	7.1
9P1N1	"Average growth"							6.6

From the 17 fast-growing SC strains (10% of screened strains), seven strains were from group A soils, six from group B and four from group C soils. For the SC strains, only one was acid-adapted (40P1N1), with the rest being able to grow at a broad pH range or in alkaline pH. Three strains (79P3N2, 6P1N1 and 78P2N1) which originated from group B soils demonstrated adaptation to a broad pH range and three strains (12P1N2, 84P2N2 and 83P1N1) showed alkaline-adaptation (Table 3.8).

Table 3.8: Rhizobia strains isolated from Subterranean Clover (SC) with faster than average growth rate in pH adjusted liquid culture, showing media pH where strains exhibited above average growth rate (presence of **+**) and the pH of soil of origin (red = Group A, green = Group B, blue = Group C). Strains 83P3N2, TA1 and WSM1325 were included for a comparative aspect. Average growth (**+/-**) and below average growth (**–**)

	pH where strains demonstrate high growth							
SC Strains	**4.5**	**5.4**	**5.8**	**6.5**	**7.5**	**8.0**	**9.0**	**Soil of origin pH**
42P1N1	+							6.1
40P1N1	+	+	+					5.5
6P2N2		+	+	+				6.1
38P3N1		+		+				5.2
79P3N2		+	+	+	+	+	+	6.0
6P1N1		+	+	+		+	+	6.1
76P2N2			+	+				5.5
82P3N1			+			+	+	6.5
73P3N2			+	+		+		4.9
78P2N1			+	+	+	+	+	6.1
4P1N1				+	+		+	5.5
43P3N2				+	+	+	+	6.0
12P1N2					+		+	7.5
76P2N1					+	+	+	5.5
84P2N2					+	+	+	7.5
83P1N1					+	+	+	7.1
38P1N1						+	+	5.2
83P3N2	"Average growth"							7.1
TA1	–	+/-	–	–	+/-	+/-	+/-	Commercial strain
WSM1325	–	+/-	+/-	+	+/-	+/-	+/-	Commercial strain

The number of pH-adapted SC and WC strains are shown as a proportion of strains cultured from nodules of plants grown in either acid, neutral or alkaline pH soils (Table 3.9).

Table 3.9: Proportions of pH-adapted SC and WC strains, calculated based on number of strains from Table 3.5.

	Subterranean Clover (SC)	**White Clover (WC)**
Acid-adapted	2%	3%
Alkaline-adapted	6%	4%
Broad pH adaptation	6%	5%

3.4 Discussion

3.4.1 Summary of findings

The development a robust bioassay contributed towards the goal of assessing the relationship between pH of soil of origin and the pH preference of rhizobia resident in those soils, testing the hypothesis that "*there is a relationship between the pH of the soil of origin and the ability of rhizobia to grow at that pH*". The results showed that the media pH where strains demonstrated fastest growth (media optimum pH) was related to the pH of the soil that strains were isolated from. However, the correlation between media pH and soil pH was strongly influenced by the growth of strains from alkaline soils (alkaline-adaptation), especially in strains isolated from SC.

3.4.2 Soil physicochemical properties

The 12 soils used for this chapter were a sub-set of the 42 used in Chapter 2. The measurements of their soil physicochemical properties were not repeated, and hence the correlations of the subset are similar to those for the complete set of soils described in Chapter 2. The observed correlations between soil pH and the other measured physicochemical properties of the soils used in this study were as expected because they exhibit a high degree of collinearity. Soil pH is closely related to levels of minerals and organic matter (see footnote 2, section 2.4.4). It is important to note that the soil tests were conducted on pooled soil samples of the top 10–15 cm from each paddock, and no differentiation was made between bulk soil and rhizosphere soil. Research has shown that there are biologically significant differences in physicochemical properties between bulk (loosely-bound or tightly-bound) soil and rhizosphere soil (Fan et al. 2018), e.g. diazotrophic communities differed between the rhizosphere and bulk soils of wheat fields, and the network of the diazotrophic community in the rhizosphere soil was less competitive and more stable compared with the bulk soil.

3.4.3 Relationship of MPN of rhizobia with soil physicochemical properties

As expected and in agreement with the literature, rhizobia abundance in soil (as determined by MPN) for both clovers had a positive correlation with soil pH, indicating that higher pH increased the abundance of rhizobia able to form nodules on SC and WC roots (Table 3.3). This result was in agreement with the nodule count results from section 2.3.4 (and references cited in Chapter 2), where the number of nodules in SC and WC increased with soil pH. The work of Wigley (2017) and Seth (2017) reported soil rhizobia populations that were similar to those in this chapter; counts of 9.2×10^2 (with lime application) and 3.3×10^2 cells/g of soil (without lime) (Wigley 2017) and 10^3–10^5 cells/g of soil from 25 soil samples (Seth 2017) of WC nodule bacteria, however they did not assess rhizobia populations in presence of a pH gradient. A survey of 98 soils collected across NZ using white clover (cv. Tribute) as the bait plant for rhizobia populations found a mean level of 4×10^4 cells/g of soil and median of 4×10^3 cells/g (Ridgway, unpublished data). These numbers are

similar to those reported overseas. For example, composite data from several studies revealed that 169 out of 236 (72%) intensely managed South Australian agricultural soils had >1000 cells of *R. leguminosarum* bv. *trifolii* per gram of soil (Howieson and Ballard 2004). Similar numbers have also been reported for other legumes species. In Brazil, Andrade *et al.* (2002) evaluated rhizobia populations in *Phaseolus vulgaris* and found a similar range (10^3 to 10^4 cells g^{-1} of soil) in soils of pH ~4.0_{CaCl2} with and without lime application.

The influence of pH-related factors on MPNs was similar to findings from section 2.3.4. The effects of K, Ca, Mg, Al and OM on MPNs are likely to be as a result of the influence of soil pH on these properties (see footnote 2, section 2.4.4). The MPN results show that the soils used in this study have a similar abundance of rhizobia over a range of soil pH to those of other studies and the data generated is therefore comparable, i.e. these soils did not exhibit any aberrant characteristics. It also showed that it was feasible to recover strains of rhizobia from plants grown in low pH soils, thus, the hypotheses could be tested across the pH spectrum.

3.4.4 Identity of bacteria recovered from nodules

Most of the strains cultured from nodules of both clovers were identified by *16S rRNA* gene sequencing as *Rhizobium* spp., with only a small percentage of cultured strains belonging to other bacterial genera. It was expected that the majority of the cultured strains would be *Rhizobium* spp., because they were isolated from nodules of clovers and this is a known host/bacterium symbiosis. These results are supported by the NGS work (section 2.3.1), where it was also observed using culture-independent methods that a majority (~90%) of nodule occupants from SC and WC roots were *Rhizobium* spp.. However, not all strains were rhizobia and this was also supported by recent work, both internationally and in New Zealand, which has shown that other bacterial genera are found in the nodules of legumes (Hartman et al. 2017; Wigley et al. 2017). Studies (reviewed in Martínez-Hidalgo and Hirsch (2017)) have shown that many of these bacteria do not fix nitrogen nor do they contain nodulation genes and their role in the legume nodule is the subject of current research (i.e. non-rhizobial nodule-inducing bacteria or non-nodulating bacteria). The non-*Rhizobium* bacteria identified here (*Bacillus*, *Herbaspirillum*, *Ochrobactrum*, *Sinorhizobium*) were also found in Chapter 2 and/or in the work by Seth (2017). However, it was noted that the four strains identified as *Ochrobactrum* were cultured only from nodules of plants grown in alkaline soils, although the significance of this was unclear given the small number of strains recovered. The 'other' genera identified here have also been described from plants and/or rhizosphere soil, e.g. *Herbaspirillum* (Olivares et al. 1996); *Neorhizobium* (Soenens and Imperial 2018); *Ochrobactrum* (Bathe et al. 2006). They may be of significance in pH adaptation (or other stress tolerance) by the host plant, but further work is required to address their functional role(s). The high representation of rhizobia species in the recovered cultures was also favored by the use of YMA to culture the bacteria, as it is optimal for rhizobia (Vincent 1970), but other bacteria are also able to grow on it.

3.4.5 Strain adaptation to pH

There was variation in the growth of the rhizobia strains in liquid culture at different pH. A majority of SC (89%) and WC (87%) strains demonstrated 'average growth' in the different pH-media (Figure 3.5). This observation fitted Connell's (1978) 'intermediate disturbance hypothesis' because only few strains grew well at pH extremes (high intensity of disturbance) and most strains grew fastest in liquid cultures at pH 5.8–7.5 (intermediate disturbance intensity). Broad range pH adaptation was evident in some strains isolated from both SC and WC. This was demonstrated by 6% (SC) and 5% (WC) of strains originating from group B (pH 5.8–6.1) soils having fastest growth in a range of pH-amended broths. In addition, four WC strains were pH-generalists, i.e. they did not originate from the group B soils, but showed fastest growth over a broad range of media pH. In contrast, SC strains which demonstrated fastest growth over a broad range of media pH, all originated from group B soils. The possible mechanisms for broad range pH adaptation are: (i) these strains had high fitness and were fast-growers, irrespective of media pH they were growing in, (ii) the strains were better at maintaining their intracellular pH (Dilworth et al. 2001), regardless of external pH, (iii) the strains were genotypically similar, particularly for the *pha* genes, which encode membrane proteins, involved in cation transport (among other functions) (Putnoky et al. 1998). Rhizobia live in multiple habitats, i.e. bulk soil, rhizosphere soil and inside plant roots. These three habitats have differences in pH, as well as availability of minerals and carbon sources, hence rhizobia may benefit from being pH-flexible in these different environments. Although pH adapted strains were identified, their specific relative abundance in soil was not measured. Broad range pH adaptation could have valuable commercial implications if these strains can persist well in a wide range of soil environments.

The increased growth of some strains over a broad pH range could have significance in their application as potential commercial inoculants. Those strains could be genotyped by GBS (Stefani et al. 2018) or whole genome sequenced (Luo et al. 2014) to assess how pH-adaptation phenotype is correlated with their genotypes. Adaptation and growth under a range of environmental pH levels requires the bacterium to access and metabolise C compounds for growth, and pH has been shown to have a strong influence on the carbon availability in soils (Andersson et al. 2000; Kemmitt et al. 2006; Rousk et al. 2009). The association between pH-adaptation and the ability of selected strains to utilise C has the potential to generate fundamental knowledge on the possible mechanisms that underpin those phenotypes. If pH-adapted strains can also demonstrate increased growth on a variety of C-sources, or on unique C-sources, strain formulations could be augmented with those C-sources in order to confer a competitive advantage when applied as inoculants. Thus, the potential of those strains as commercial inoculants would increase.

Overall, there was a low percentage of strains exhibiting adaptation to pH extremes, but the proportion of acid-adapted strains was lowest (Table 3.9). With the exception of one SC and one WC strain, strains which originated from group A soils were not pH-specialists, i.e. their high-

growth was not restricted to acidic broths (Table 3.7 and 3.8). The reasons for observing a lower frequency of acid-adaptation are likely to include: (i) Evolving acid-adaptation may be energetically costly, because strains also have to cope with metal toxicities as soil pH drops (higher levels of Al^{3+}, Mn^{2+}, Zn^{3+}) (Foy 1984; Dilworth et al. 2001), (ii) there is a shorter history of the sampled soils being acidic (information from Landcorp farm managers was limited to a five-year historical record). With the exception of hill country soils in NZ, very few pastoral soils remain naturally acidic because they are heavily managed (i.e. lime and fertiliser inputs) (Lowther and Kerr 2011) and this may decrease selective pressure for strains to become acid-adapted, (iii) clover does not grow well in acidic soils (Andrew 1978; Maxwell et al. 2012), hence low population size of the host plant would correspondingly decrease the population of rhizobia strains and thus lower the frequency of a strain adapting to a low pH environment and (iv) there is a lower abundance of rhizobia (MPN) present in acidic soils.

The proportion of strains (from both clovers) demonstrating alkaline-adaptation (i.e. strains which originated from group C soils and grew more in broth of pH > 7.0) was greater than those exhibiting acid-adaptation (Table 3.9). Alkaline-adaptation was more evident in SC strains than WC strains. The fast-growing SC strains which originated from group C soils generally had increased growth in broths of pH > 5.8 and could not grow well in low pH broths. The correlation between media optimum pH and soil pH was not significant when data from strains of both SC and WC originating from soils of pH 7.1 and 7.5 were removed. This suggested that strains from group C soils were driving the correlation. Two of the higher pH (7.1 and 7.5) soils sampled and used for baiting rhizobia had a long-standing history of alkalinity. Those soils were sampled from Napier (Hawkes Bay) from paddocks which were upheaved from the ocean due to the 1931 earthquake. Since, those paddocks had maintained alkalinity for 87 years, the sustained selection pressure on introduced rhizobia strains may have driven adaptation to an alkaline environment. However, it could also simply be that alkaline adaptation (at least for pH 7.0–9.0) is potentially easier to achieve since most bacteria maintain their intracellular pH at ~7.8 (Booth 1985). Previous work on mutants and wild-type *S. meliloti* strains of the *pha* genes suggested that K^+ and/or Na^+ transport (efflux) systems may be involved in alkaline adaptation (Putnoky et al. 1998).

3.4.6 Development and effectiveness of the liquid culture pH assay

This work detailed the development of a robust and high throughput *in vitro* assay for growth of rhizobia strains over a range of pH. This *in vitro* liquid culture assay enabled growth of strains to be evaluated with only one environmental variable - pH, without being influenced by confounding factors such as temperature, salinity, mineral availability and competition from other bacteria/strains. There are examples of microbial bioassays on phosphate solubilisation (Barea et al. 1976; De Freitas et al. 1997; Nautiyal 1999) and soil contamination (Keddy et al. 1995;

Ronnpagel et al. 1995; Chapman et al. 2013), but no study has comprehensively tested growth of rhizobia strains on pH-amended broth spanning a wide pH range.

The advantages of performing *in vitro* assays is that strains from different plant hosts can be used together and hence comparisons are easier to make as experimental variation is limited. Much work has been conducted to screen phosphate solubilising bacteria both quantitatively (Nautiyal 1999) and qualitatively (Mehta and Nautiyal 2001) using *in vitro* plate (Seth 2017) or broth (Johri et al. 1999) assays. In addition, rhizosphere bacteria have been screened for plant growth promoting traits using indole acetic acid (IAA) assays, ammonia assays and siderophore production assays (Ahmad et al. 2008). These *in vitro* assays were useful for screening numerous (4800) bacterial isolates (Johri et al. 1999), from different plant hosts (SC and WC) (Seth 2017) and/or their ability to produce a range of compounds (Ahmad et al. 2008).

Another benefit of this pH assay is the low cost, especially when compared with commercial plates (e.g. Biolog™ PM10) which include pH-adjusted media with or without nutrient amendments. The pH of the liquid culture at the end of the growth period for each of the six strains was similar, making relative growth comparisons possible across all strains and all pH-amended broths. Using pH buffers may have negated an important adaptive mechanism for the rhizobia, as being able to acidify the growth medium may have enhanced their growth and survival. Using buffers may therefore hinder the expression of such mechanisms which could be intrinsically linked with pH tolerance and/or pH homeostasis (Padan et al. 2005; Krulwich et al. 2011). However, it is possible that the rhizobia were demonstrating more fundamental biochemical mechanisms while growing in the pH-amended media rather than specific pH adaptation mechanisms.

This method only tested growth of strains in broth, and although differences in growth were observed in the different pH-amended broths, it may only be because some strains grew better in liquid culture than others. Furthermore, only YMA was empirically used to assess strain performance. Based on the robustness of the study and number of replicates used, an assumption could be made that relative strain performance may not change significantly if different growth media were used. However, this is only an assumption because different media were not tested. If a 'media effect' did exist, i.e. strain performance was altered based on different growth media, irrespective of media pH, then this points to nutrient differences in the media rather than specific pH adaptation/tolerance mechanisms. Another disadvantage of using this method to select pH-tolerant/adapted strains is that the broth did not take into account concentrations of metal ions which are important in a soil context, such as Ca^{2+}, Al^{3+}, or Zn^{3+} (Dilworth et al. 2001). As with other functional tests, such as phosphate solubilisation, *in vitro* assays provide a primary screening system and subsequent validation of candidate strains within the specific field environment is necessary. However, the initial laboratory tests are a quick and

easy way to eliminate a majority of strains which do not demonstrate field-potential (Dilworth et al. 2001).

Future work could test the identified divergent (fast-growing) strains for their symbiotic potential to determine whether they have potential as commercial inoculants (Ballard et al. 2002; Ballard et al. 2003). This would involve inoculating clover seed with a coat containing selected strains and assessing SDW (or other yield-related parameters) in the field. Those strains that are acid tolerant may be particularly interesting as they could enhance clover establishment and nitrogen delivery to the New Zealand high country. Preliminary *in vitro* work (data not shown) indicated that the symbiotic potential of most of the selected SC and WC strains when grown in sterile vermiculite was not significantly different to the relevant commercial strains WSM1325 and TA1 (Table A 18, Chapter 3 supplementary results). Validation under field conditions would show whether some pH-adapted strains are able to increase SDW in clover under a challenging soil pH, then those strains could have potential as new commercial inoculants for SC and/or WC in acidic soils.

Rhizobia strains from the culture collection can then be tested for (functional) traits which make them specialised at surviving and persisting in either a narrow or broad pH range. This has potential to provide valuable information about rhizobial ecology, both from an applied perspective as well as to advance our scientific knowledge.

3.5 Conclusions

- A robust and efficient assay was developed to screen strains of rhizobia for pH-adaptation/tolerance.
- Relatively few strains demonstrated preference for a narrow range of pH conditions, and the majority grew best in broths of pH 5.8 – 7.5 (which corresponded to predictions of the 'intermediate disturbance hypothesis').
- For the first time, there was evidence of alkaline-adaptation by SC and WC strains isolated from soils in group C, but acid-adaptation was not specifically associated with strains recovered from soils in group A.
- Some SC and WC strains isolated from group B soils demonstrated faster than average growth over a broad pH range.
- Rhizobia populations (MPN) in soil were positively correlated with soil pH (and related properties), confirming previous reports.
- A majority of bacteria cultured from nodules of both clover species were identified as *Rhizobium* spp., which is consistent with results from Chapter 2.

In the following chapter (Chapter 4), the divergent strains of WC identified from the pH assay will be screened to assess their carbon utilisation profiles. This may provide information on whether pH-adaptation in some strains can be linked to a competitive advantage in soils based on

metabolism of a diverse range of carbon sources. The growth of strains in the pH assay was supported by their ability to metabolise mannitol as the primary carbon source. However, for commercial application, it is important to test the ability of rhizobia strains to utilise a variety of other carbon sources, such as those found in bulk or rhizosphere soils (for free-living rhizobia) and those supplied by the plant (for nodule-dwelling rhizobia).

4 Carbon utilisation by pH-adapted rhizobia strains isolated from white clover nodules

4.1 Introduction

In Chapter 3 the results showed that, although rare, some strains of rhizobia exhibited adaptation to pH. A correlation existed between the ability of a strain to grow in liquid media at a particular pH and the pH of the soil of origin, however, this was strongly influenced by strains derived from soils with pH >7.0. Rhizobia typically contain a large multipartite genome composed of a chromosome and plasmids (Young et al. 2006). These core and ancillary (plasmid) genomes may be ecologically advantageous in the diverse and/or challenging soil environments (Mazur et al. 2013; Stasiak et al. 2014). The core genome for *R. leguminosarum trifolii* biovars is approximately 4.8 million base pairs (bp) in size, however the ancillary genome (often consisting of five or more large plasmids), can be variable and constitute ~32% of the total bacterial genome size (Terpolilli et al. 2014; Delestre et al. 2015). The movement of plasmids among strains in soil populations of rhizobia may improve tolerance of a range of soil physicochemical properties. Studies (reviewed in Lodwig and Poole, 2003) have shown that the transport and catabolism of dicarboxylic acids (such as malate, fumarate, succinate, aspartate) by *R. leguminosarum* and *S. meliloti* are important for the reduction of atmospheric nitrogen into ammonia inside the nodules (Lodwig and Poole 2003). Wielbo *et al.* (2007) demonstrated using Biolog GN2 microplates™ that strains of *R. leguminosarum* bv. *trifolii*, *R. leguminosarum* bv. *viciae*, and *R. etli* utilised different carbon sources. The authors found that the ability of rhizobia strains to metabolise a broad range of substrates and their nodulation competitiveness was significantly correlated, but there have been no studies comparing pH-adaptation and C-utilisation of strains.

The different habitats occupied by rhizobia (bulk soil, rhizosphere soil and root nodules) differ in their physicochemical properties, and also in the type of carbon sources present. Several studies have found quantitative and qualitative differences in carbon sources in the different niches occupied by rhizobia. Root exudates are one of the main sources of organic carbon in the rhizosphere (Lynch and Whipps 1990; Bertin et al. 2003; Kumar et al. 2006) and in turn account for the higher carbon availability in rhizosphere compared with bulk soil (Cheng et al. 1996). Given that rhizobia need to occupy all these habitats, a diverse repertoire of metabolic pathways is required to effectively utilise the various resources available. To support these metabolic pathways is energetically costly for rhizobia (Slater et al. 2008), thus a method by which a diverse metabolic potential is held in the population is the mobile genome, i.e. the plasmids.

The strains derived from WC were the subject of this chapter due to the high value of this legume to New Zealand pastoral systems (Caradus et al. 1996). Studying the metabolic diversity of strains that nodulate white clover and that have been specifically selected based on their ability to grow

in a narrow or broad pH-range, will contribute to understanding whether there is a relationship between pH adaptation and ability to utilise various C sources. Previous research on carbon transport and/or metabolism in rhizobia have been conducted on strains of *R. leguminosarum* bv. *viciae* (pea symbiont) and a single study on *R. leguminosarum* bv. *trifolii* that nodulates red clover (*T. pratense*) (Wielbo et al. 2007; Wielbo et al. 2010). The goal of this chapter was to determine whether the rare strains of rhizobia that demonstrated adaptation to specific or broad pH ranges could be differentiated by their ability to utilise particular C sources. To achieve this goal the following objectives were developed: 1) To phylogenetically identify the pH-adapted strains (from Chapter 3) based on *16S rRNA* and *nodC* genes, 2) To characterise the carbon utilisation profile of pH-adapted strains and 3) To determine whether C utilisation correlated with pH adaptation.

4.2 Materials and Methods

4.2.1 *16S rRNA* and *nodC* sequencing of selected pH-adapted strains

The 19 strains used for metabolic profiling were sent to Macrogen Inc. (South Korea) to be sequenced for the *16S rRNA* gene. Their facility performed the DNA extractions, PCRs with F27 and R1492 primers, Sanger-based sequencing of PCR products using F785 and R907 primers, assembly of the consensus sequences and provided the full reports for the service. These sequences are available on GenBank under accessions MH915594 through to MH915608.

The 19 strains were also sequenced for the symbiotic *nodC* gene to determine if there were genetic differences between pH-adapted strains. The methods to obtain single colonies and generation of a PCR master mix were as described in section 3.2.4. The protocols used to amplify products of each gene are outlined in Appendix 2.

The *nodC* PCR products were separated by electrophoresis on a 1% agarose gel in 1× TAE buffer at 100 V for 45 min. Bands of the appropriate size and concentration were directly sequenced in both directions at the Lincoln University Sequencing Facility. The sequences obtained were viewed and trimmed using Geneious (v 11.1.4) to remove ambiguous sequence ends and then assembled (by Aurélie Laugraud). The consensus sequences for *16S rRNA* were ~1,400 base pairs (bp) long and those for *nodC* were 500–600 bp long. The phylogenetic identity was determined by comparing sequences to those of known origin using the nucleotide basic local alignment search tool (BLASTn) against deposited sequences in the NCBI GenBank database (www.ncbi.nlm.nih.gov/genbank/). Phylogenetic (maximum likelihood) trees for the genes were built by Aurélie Laugraud in Geneious (v 11.1.4) using the PhyML algorithm (Guindon et al. 2010) with a general nucleotide substitution model (GTR model) and 512 bootstraps. The trees were rooted to known bacterial sequences (from NCBI GenBank database) for the specific genes.

For *16S rRNA*, genetic distances (evolutionary divergence) between pairs of samples were calculated using the Maximum Composite Likelihood model (Tamura et al. 2004). All positions containing gaps were eliminated.

Permutation based multivariate analysis of variation (PERMANOVA; (Anderson et al. 2008)) was used to determine if *16S rRNA* genetic distances were related to generalised soil order (Table 4.1). BIO-ENV matching (Clarke and Ainsworth 1993; Clarke 1993) was used to determine links between genetic distances and edaphic and climatic properties.

The RELATE test (Clarke 1993) was used to determine whether genetic distance was associated with geographic distance. A resemblance matrix was generated based on the geographic distance

(Euclidean distance) and RELATE was used to test for correlation (ρ) to the genetic distance matrix. Statistical significance was determined using permutation-based testing (× 999).

4.2.2 Carbon utilisation assay

The results from the pH-bioassay were used to select strains which demonstrated divergent growth in pH-media for characterisation using Phenotype MicroArray™ (PM) plates (Bochner et al. 2001). The OmniLog® PM system (Biolog Inc., Hayward, CA) was used to characterise 19 WC strains and two commercial strains (TA1 and WSM1325) on two types of carbon source plates (PM01 and PM02, Biolog™). Each plate consists of 95 individual carbon sources and one water control (96 wells), allowing the growth of selected isolates to be tested across a total of 190 carbon sources. The carbon sources are placed in the following groups (Biolog™): alcohol, amide, amine, amino acid, carbohydrate, carboxylic acid, ester, fatty acid and polymer.

Strains were grown by spreading 120 µL of liquid culture on TY media and incubating at 25°C for five days, or until a reasonable (~80% area coverage) lawn of bacterial culture was visible on the plates. The Inoculating Fluid (IF-0) was prepared at 1× concentration by adding 1.40 g of supplied IF-0 (Biolog Inc.) to 1 L of H_2O, of which 15 mL was transferred into sterile, capped tubes. Swabs of the bacterial lawn were suspended in the fluid and stirred to achieve a uniform suspension. The turbidity of the suspension was measured in the Biolog™ Turbidimeter to confirm it achieved a 15% transmittance (T). For both PM01 and PM02 plates, 1.76 mL of the cell suspension (15% T) was added to 2 mL of sterile distilled water, 240 µL Redox Dye mix G and 20 mL of 1.2× IF-0 (1.68 gL^{-1} of H_2O). This final cell concentration produced a transmittance of 85%.

The PM01 and PM02 plates were then inoculated with this cell suspension at 100 µL per well. The plates were loaded into the Omnilog incubator and incubated at 28°C for 54 h. During this time, automatic optical density readings were taken every 15 min. Kinetic and parametric analysis was performed using the software supplied by Biolog™ and data parameters for various growth metrics were exported for statistical analyses.

The experiment was split into four runs, with each run having six strains and two replicates for each PM plate. One strain (74P3N2, which demonstrated growth in a broad pH range) was used as an internal standard and was included in all four runs. One of the replicates from run 1 had a problem whereby the strains failed to grow (reason unknown) and so data from run 1 were not replicated.

4.2.3 Climatic and geographic data

GPS coordinates were used to determine broad soil order (Hewitt 1998) within the New Zealand Fundamental Soils Layer of the landcover database (Newsome and Pairman 2012). The strain

names (as described in Chapter 3, section 3.2.2) and corresponding sampling information are shown in Table 4.1 (for more detailed soil data refer to Chapter 2 supplementary results). Geographic distances among sample points were calculated using the Geographic Distance Matrix Generator (http://biodiversityinformatics.amnh.org/open_source/gdmg/). Historic environmental data were interpolated for each point based on the virtual climate station network. These were used to determine average daily maximum and minimum temperatures (°C), rainfall (mm over 24 h), solar radiation (MJm^{-2} over 24 h), soil moisture (kgm^{-2} over 24 h) and Penman evaporation ($kgm^{-2}s^{-1}$ over 24 h) averaged for the 2014–2015 period. Summary climate data are shown in Chapter 4 supplementary results.

Table 4.1: Selected strains, locations they were isolated from, associated GPS coordinates, their broad soil order and soil pH values.

Strain	Location	Latitude	Longitude	Soil order	Soil pH	pH group
1P1N1	G13+BH29	-41.79609	171.56245	Podzol	4.9	Group A
73P1N2	G13+BH29	-41.79609	171.56245	Podzol	4.9	
74P3N2	Bills	-38.85353	177.13721	Pumice	5.2	
4P1N1	AR_120	-38.62847	176.15197	Pumice	5.5	
41P1N2	TO_39	-41.79182	171.50091	Podzol	5.8	Group B
41P3N2	TO_39	-41.79182	171.50091	Podzol	5.8	
5P1N2	TO_39	-41.79182	171.50091	Podzol	5.8	
5P2N1	TO_39	-41.79182	171.50091	Podzol	5.8	
5P2N2	TO_39	-41.79182	171.50091	Podzol	5.8	
43P2N2	DG_D16	-45.62706	167.66432	Brown	6.0	
79P2N2	DG_D16	-45.62706	167.66432	Brown	6.0	
6P1N2	TO_C06	-41.77501	171.49853	Podzol	6.1	
78P2N2	TO_C06	-41.77501	171.49853	Podzol	6.1	
80P3N2	CH_F38	-45.52276	168.12373	Brown	6.1	
46P2N1	SU_37	-42.49942	171.59457	Podzol	6.5	Group C
81P2N2	ER_38	-38.58004	176.25165	Pumice	6.6	
9P1N1	ER_38	-38.58004	176.25165	Pumice	6.6	
47P2N1	SiberiaWatch	-39.47264	176.84986	Recent	7.1	
84P2N2	Pidgeon13	-39.47182	176.85151	Recent	7.5	

4.2.3.1 Statistical analysis

The Area Under the Curve (AUC) metric for was used for the data analysis. A minimum threshold was also set in order to determine which carbon sources substantially promoted growth in the strains. To do this, only carbon sources with AUC values (averaged across all strains) which were ≥ 1.2 × (AUC of Negative control) for both PM01 and PM02 plates were used, i.e. at least 20% growth promotion over control. AUC values less than the threshold were deemed to be within the range of biological variability. This method resulted in the removal of 69 carbon sources (24 from PM01 and 45 from PM02) from further analyses (Table A 23).

For the remaining 121 carbon sources (Table A 19), plus two negative controls, the data from the four runs were standardised using the internal standard (strain 74P3N2). These data were viewed and analysed using R v3.4.3 (Team 2017) to elucidate differences between growth of strains and differences between utilisation of the different types of carbon sources.

For the multivariate analyses, the AUC data were standardised for each strain (excluding the commercial strains) in PRIMER v7 (Clarke and Gorley 2015). Hence, effects of inherently fast or slower growing rates among strains were removed, allowing for the underlying responses to the patterns of C source utilisation to be expressed. The standardised AUC data were square-root transformed and similarity in C-utilisation among strains determined using the Bray-Curtis method.

Cluster dendrograms were plotted using the group average method to view natural SIMPROF groups, i.e. without the influence of environmental factors, allowing for the identification of phenotype groups at $\alpha = 0.05$. A SIMPER analysis was conducted to view the influence of specific carbon sources on those phenotype groupings.

PERMANOVA was used to formally test for the presence of treatment structure within the Bray-Curtis calculated AUC distance matrix. Tests (one-way) were conducted for 'location', and 'soil order' treatment. Statistical testing was conducted using unrestricted permutations (999) of the raw data to create a null distribution against which the treatments could be assessed.

Links between the strain phenotypes and underlying soil physicochemical properties (n = 17) were tested using BIO-ENV matching. This method identified combinations of soil variables that maximised the Spearman rank correlation (ρ) to the fixed phenotype distance matrix. Permutation of variables (199 permutations) was used to generate a null-distribution to statistically allow probability testing. A maximum of five variables were used in combination.

Given previously reported associations between soil pH and microbial community structure and function (Wakelin et al. 2008; Lauber et al. 2009), an *a priori* hypothesis was raised, that pH of the soil of origin is linked with strain phenotype. The RELATE test was used (as before) to determine correlation between soil pH and strain phenotypes. The RELATE test was also used to determine correlation between similarity in phenotypes with (1) genetic distance among samples, (2) geographic distances among samples, and (3) climatic conditions. Similarly, BIO-ENV matching was used to explore if specific climatic variables, or any combinations of these, were related to phenotypes.

All multivariate statistical testing was conducted in PRIMER v7 (Clarke and Gorley 2015) and PERMANOVA+ (Anderson et al. 2008).

4.3 Results

4.3.1 *16S rRNA* and *nodC* sequencing results

Neighbour Joining (NJ) trees for *16S rRNA* (Figure 4.1) and *nodC* (Figure 4.2) genes are displayed below. The strains are labelled according to the pH of the soil of origin. The *16S rRNA* NJ tree showed that strains grouped into five genotypes: one group with strains 84P2N2, 79P2N2 and 47P2N1, three separate groups with one strain each (74P3N2, 73P1N2 and 81P2N2) and one group with the remaining 13 strains.

There was no variation in *16S rRNA* genotypes among sites with generalised soil order (PERMANOVA p = 0.501, √CV = -0.006). However, the level of replication of some soil orders was low; as such, this finding requires further validation. The best combinations of soil properties correlated with *16S rRNA* genotype were soil pH and either Na or Olsen P (ρ ~ 0.420). However, these multi-variable associations were not significant (p = 0.275). There was a trend (ρ = 0.227; p = 0.067) to support a relationship between *16S rRNA* genetic distance and the geographic (sampling) distance. However, there was no significant correlation of any of the climatic properties with the genetic distance between strains (ρ = 0.334; p = 0.105).

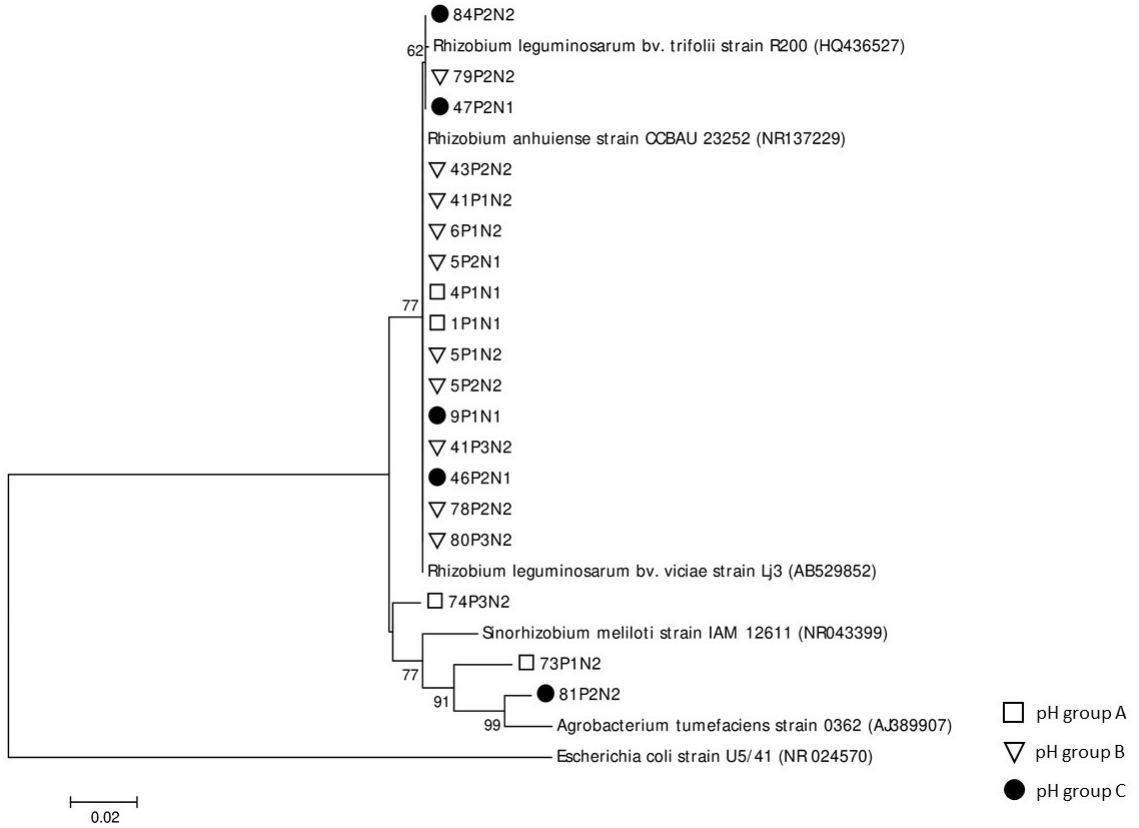

Figure 4.1: Phylogenetic tree of 19 strains for the *16S rRNA* gene with reference strains from the NCBI database and the corresponding accession numbers in parentheses. Bootstrap values are shown as a percentage (> 50%) and strains are annotated with symbols corresponding to their soil pH groups.

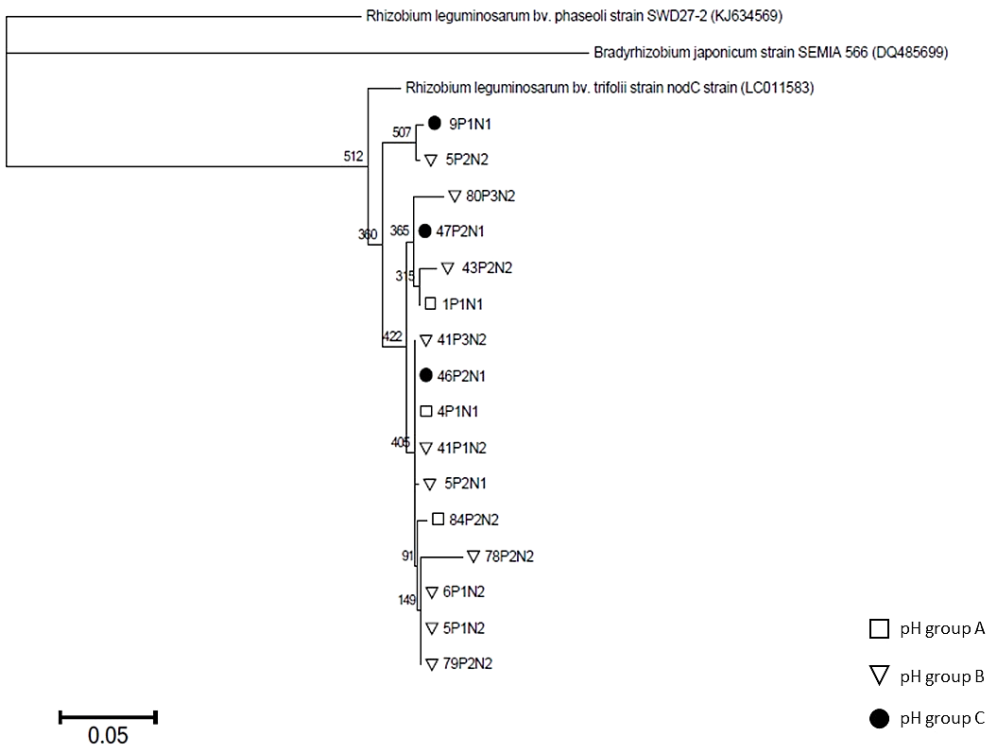

Figure 4.2: Phylogenetic tree of 16 strains for the *nodC* gene with reference strains from the NCBI database and corresponding accession numbers in parentheses. Bootstrap values are shown beside the nodes. Strains are annotated with symbols corresponding to their soil pH groups.

On searching the NCBI database, the *16S rRNA* sequence of strain 73P1N2 was found to be 99% identical to *Sinorhizobium* spp. (Accession KT724704.1); the sequence of strain 74P3N2 was found to be 99% identical to *Rhizobium* spp. (Accession DQ337551.1) and the sequence of strain 81P2N2 was found to be 99% identical to *Agrobacterium* spp. (Accession KC213941.1). Strains 73P1N2, 74P3N2 and 81P2N2 were unable to form nodules on white clover roots, when inoculated in isolation to test their symbiotic potential (data not shown). This explains why *nodC* sequences were not obtained for these three strains.

4.3.2 Carbon metabolism

Overall, the growth of each of the strains differed with carbon sources (Figure 4.3). Some strains could metabolise particular carbon sources better than other strains, but there were also certain carbon sources that were not utilised by a majority of the strains. All carbon sources were placed into nine groups (Biolog®).

A minimum threshold was set for the AUC values to identify carbon sources that substantially promoted growth rather than simply showing minor effects. This threshold removed 69 of the 190 carbon sources that had AUC values < 1.2 × (AUC of Negative control) for all replicates of all strains. The carbon sources that were removed from further analysis are listed in Table A 23 in Chapter 4 supplementary results.

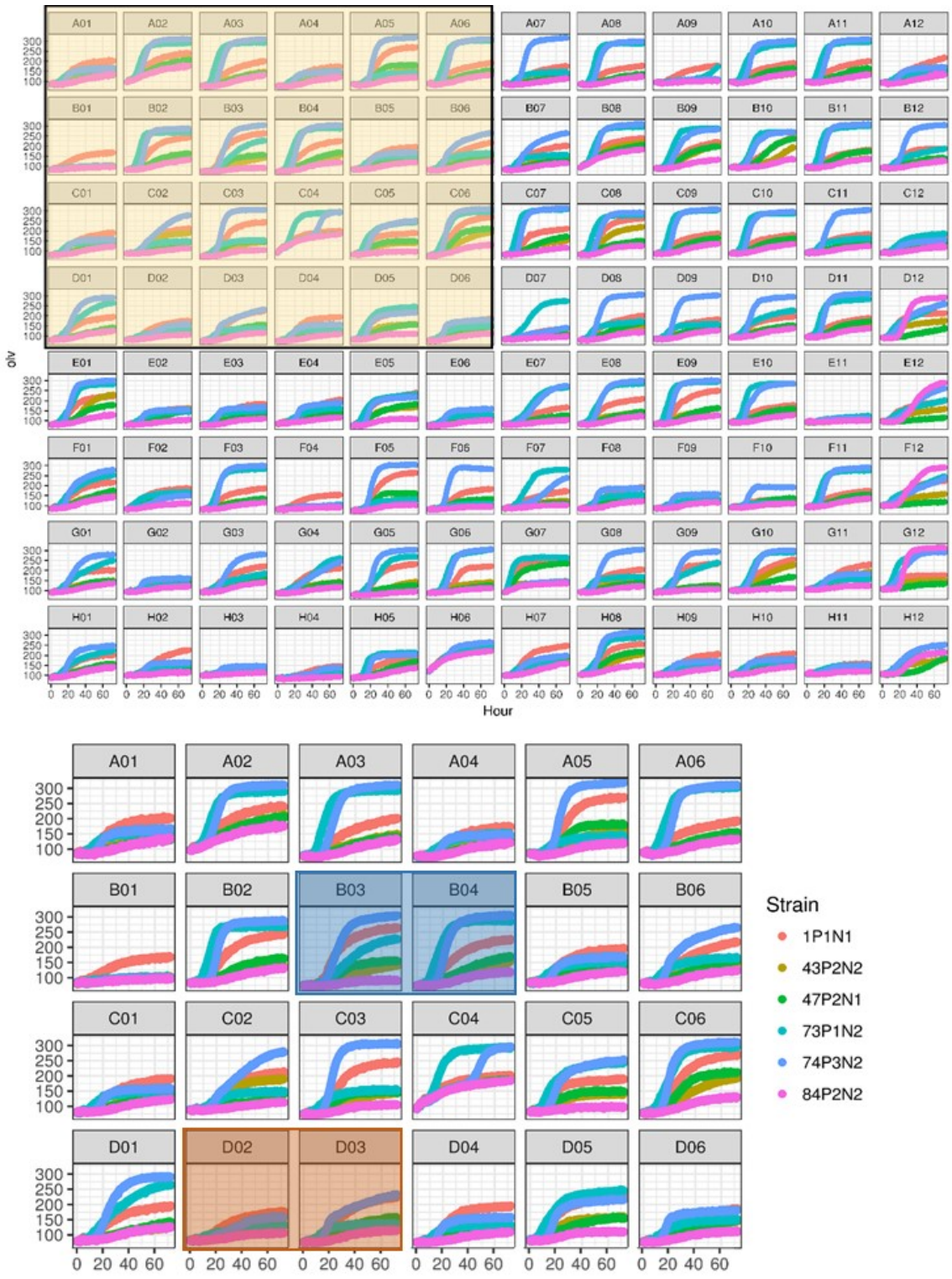

Figure 4.3: A sample of the kinetic data showing variation of carbon utilisation by six strains from 'run 2' of the experiment, grown on PM01 plates. Curves show optical density readings (y-axis) against time (x-axis). Each block is a carbon source, with A01 being the negative control. Each coloured line is a specific strain. The section in the yellow shaded box in the top image is enlarged in the bottom image to highlight variation in carbon utilisation between strains. The blue shaded box (wells B03 and B04) show examples of carbon sources that met the minimum AUC threshold, whereas the orange shaded box (wells D02 and D03) show examples that were below minimum AUC threshold.

Some carbon groups were less preferred by the strains than others (Table 4.2), since more than 50% of the total carbon sources of those groups (values in bold) produced AUCs which were lower than the set threshold. However, it should be noted that the total numbers of substrates in some groups were relatively low (e.g. only 2 carbon sources for esters). For these groups, reliable estimates of percentage of utilisation were not achievable.

Table 4.2: Number of carbon sources of the different groups that were retained or removed from the analysis and the removed carbon sources as a percentage of the total for each group.

Carbon group	Retained C-sources	Removed C-sources	% Removed
Alcohol	2	4	**67**
Amide	2	1	33
Amine	1	4	**80**
Amino acid	17	13	43
Carbohydrate	63	8	11
Carboxylic acid	26	33	**56**
Ester	1	1	50
Fatty acid	3	0	0
Polymer	6	5	45

Strains showed differences in metabolism of the carbon group (Table 4.3). The t-test results indicated that, compared with the negative control, strains grew significantly more in alcohols, amides, carbohydrates, carboxylic acids, fatty acids and polymer carbons. Negative controls and amines produced least growth (lowest mean AUCs) in the strains.

Table 4.3: T-test with the Negative controls as the intercept (reference). Values are for Area Under the Curve (AUC). p values with * are significantly different from the negative controls at 95%.

Carbon group	Diff. from intercept	Estimated mean AUC	Std. error	t value	p (>|t|)
Neg. control	0	28026	NA	NA	NA
Alcohol	6557	34583	2113	3.10	0.0019 *
Amide	4352	32378	2113	2.06	0.0395 *
Amine	-491	27535	2588	-0.19	0.8494
Amino acid	2964	30990	1575	1.88	0.0599
Carbohydrate	6332	34358	1518	4.17	< 0.0001 *
Carboxylic acid	3971	31997	1540	2.58	0.0100 *
Ester	2476	30502	2113	1.17	0.2413
Fatty acid	4295	32321	1929	2.23	0.0260 *
Polymer	7449	35475	1725	4.32	< 0.0001 *

Strains showed significant differences in their AUCs across carbon sources (Table 4.4). Strains 46P2N1, 4P1N1, 74P3N2 and 9P1N1 (blue) had significantly more growth (larger AUC) than most other strains, whereas strains 47P2N1, 5P1N2, 78P2N2, 80P3N2, 81P2N2, 84P2N2 and WSM1325 (red) had significantly less growth (smaller AUC) than most other strains.

Table 4.4: T-test with strain TA1 as the intercept (reference strain). Values are for Area Under the Curve (AUC). p values with * are significantly different from TA1 at 95%.

Strain	Diff. from TA1	Estimated mean AUC	Std. error	t value	p (>|t|)
TA1	0	33523	NA	NA	NA
1P1N1	1865	35388	577	3.23	0.0012 *
41P1N2	5176	38699	577	8.97	< 0.0001 *
41P3N2	4911	38435	577	8.51	< 0.0001 *
43P2N2	2143	35666	577	3.71	0.0002 *
46P2N1	11842	45365	707	16.75	< 0.0001 *
47P2N1	-5411	28112	577	-9.38	< 0.0001 *
4P1N1	11471	44994	707	16.23	< 0.0001 *
5P1N2	-4443	29080	707	-6.29	< 0.0001 *
5P2N1	6643	40166	577	11.51	< 0.0001 *
5P2N2	5070	38593	577	8.78	< 0.0001 *
6P1N2	5752	39275	707	8.14	< 0.0001 *
73P1N2	1995	35518	577	3.46	0.0005 *
74P3N2	8674	42197	463	18.74	< 0.0001 *
78P2N2	-3063	30460	577	-5.31	< 0.0001 *
79P2N2	7419	40942	577	12.85	< 0.0001 *
80P3N2	-10005	23518	577	-17.33	< 0.0001 *
81P2N2	-4178	29346	577	-7.24	< 0.0001 *
84P2N2	-4327	29196	577	-7.50	< 0.0001 *
9P1N1	14004	47527	707	19.81	< 0.0001 *
WSM1325	-5229	28294	577	-9.06	< 0.0001 *

To determine if there was phenotypic variation in the strains for C-utilisation, a cluster analysis was performed. From the 19 strains tested, 10 significantly different ($p < 0.05$) phenotypes were identified; these were arbitrarily labelled phenotypes ‘A’ through to ‘J’ (Figure 4.4).

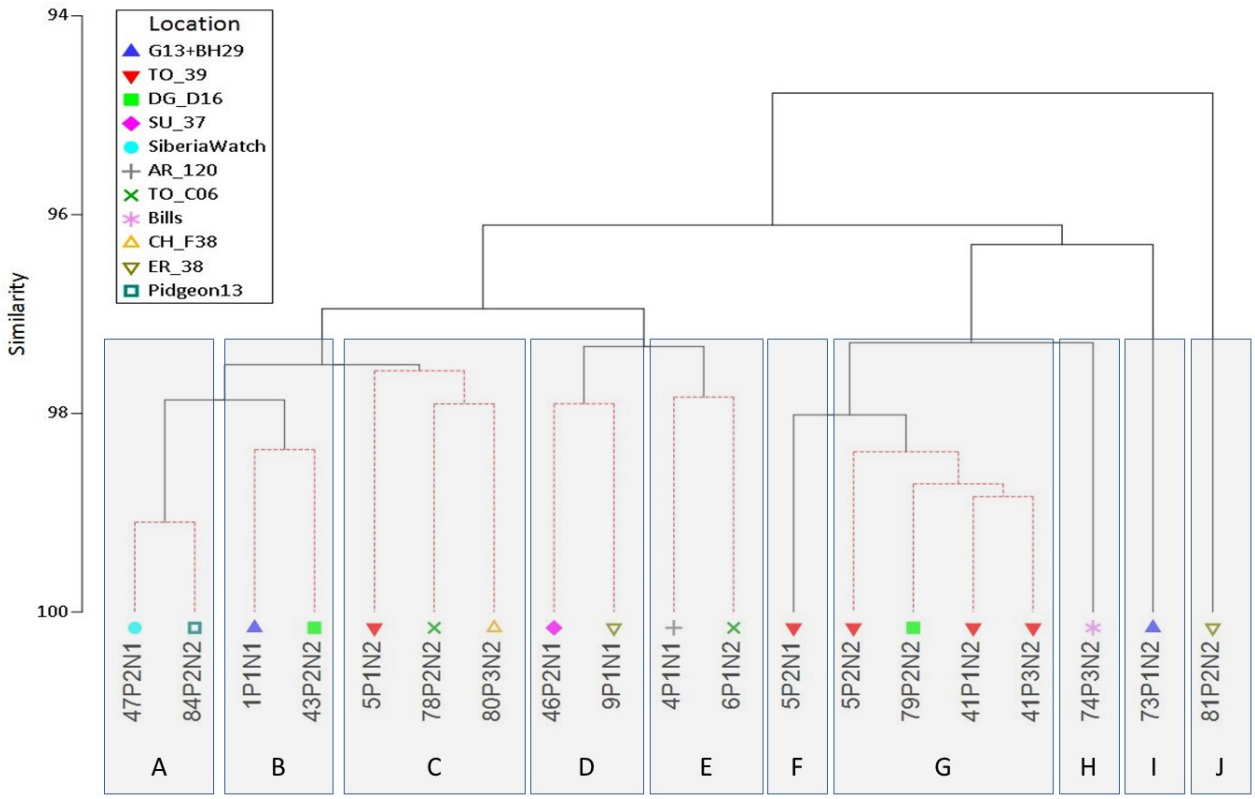

Figure 4.4: Clustering of the 19 strains by percentage similarity (Y-axis) of C-substrate utilisation (group average method). Statistically similar phenotypes are grouped by shaded boxes and designated 'A' through to 'J' on the X-axis. Location is the field of origin of the strains (refer to Table 4.1).

The ten C-utilisation phenotypes of the 19 strains had no association with the ability of the strains to grow optimally on low, medium, high or broad range of pH broths (Table 4.5).

Table 4.5: Comparison of C-utilisation groups with pH-adaptation of the 19 WC strains.

Strain	**C-utilisation SIMPROF group**	**Optimal broth pH (from Table 3.7)**
47P2N1	A	9.0
84P2N2	A	5.4
1P1N1	B	4.5
43P2N2	B	5.4 – 8.0
5P1N2	C	4.5 – 6.5
78P2N2	C	5.8 – 6.5
80P3N2	C	5.8 – 9.0
46P2N1	D	5.4 – 9.0
9P1N1	D	"Standard strain"
4P1N1	E	5.4 – 9.0
6P1N2	E	4.5 – 6.5
5P2N1	F	5.8

41P1N2	G	5.8
41P3N2	G	4.5
5P2N2	G	5.4 – 8.0
79P2N2	G	4.5
74P3N2	H	4.5 – 9.0
73P1N2	I	5.4 – 9.0
81P2N2	J	7.5 – 9.0

A SIMPER analysis was conducted to determine which carbon sources defined the different phenotypic groups, i.e. were the groupings based on utilisation of specific carbon sources by strains? Only carbon sources which contributed >5% were chosen for each SIMPROF group (Table 4.6). The analysis cannot be performed on groups which had fewer than two strains (i.e. groups F and H-J).

Table 4.6: Carbon sources which defined the different SIMPROF groups.

SIMRPOF group	**Carbon source**	**Carbon group**
Group A	Acetoacetic acid a-Methyl-D-Mannoside Tween 20	Carboxylic acid Carbohydrate Fatty acid
Group B	L-Pyroglutamic acid	Amino acid
Group C	Laminarin Hydroxy-L-Proline	Polymer Amino acid
Group D	L-Asparagine p-Hydroxy Phenyl Acetic Acid	Amino acid Carboxylic acid
Group E	a-D-Glucose D,L-Octopamine i-Erythritol L-Pyroglutamic acid	Carbohydrate Amine Carbohydrate Amino acid
Group F	None	-
Group G	L-Asparagine D-Glucose-6-Phosphate D,L-Octopamine p-Hydroxy Phenyl Acetic Acid L-Pyroglutamic acid L-Aspartic Acid	Amino acid Carbohydrate Amine Carboxylic acid Amino acid Amino acid
Group H	None	-
Group I	None	-
Group J	None	-

There was a strong relationship between the five *16S rRNA* genotypes (genetic distance) and strain phenotypic variation (ρ = 0.473; p = 0.002). Furthermore, PERMANOVA testing revealed that neither soil order (p = 0.366; √CV = 0.39) nor location of sampling (p = 0.428; √CV = 0.41) were associated with phenotype similarity.

BIO-ENV matching was used to identify if soil physicochemical variables, or combinations of these (up to 5) could explain phenotypic variation in strains. The highest single correlation was with Olsen P (ρ = 0.176); this correlation did not significantly increase with the addition of further variables (maximum ρ = 0.204 with inclusion of Olsen P, K, Al, and total C). These correlations are relatively weak and had a strong likelihood of occurring by chance (p = 0.59). In addition, the association between phenotype and soil pH was independently tested and RELATE-based analysis found no evidence for an association (ρ = 0.031; p = 0.37).

Geographic distance between samples ranged from very close (i.e. same field) to over 1,000 km apart. There was no association between geographic (sampling) distance and strain phenotype, when tested using non-parametric correlation (ρ = 0.037; p = 0.36). There were no significant associations among climatic variables, either individually or in combinations with strain phenotypes (BIO-ENV permutation derived p = 0.75).

4.4 Discussion

4.4.1 Summary of findings

This work profiled 19 WC strains (pre-screened on their ability to grow well at either low, high or broad pH range) based on C-utilisation of 190 C-sources. A high degree of phenotypic variation in carbon source utilisation was observed, i.e. 10 groups among 19 strains. This phenotypic variation was strongly associated with the *16S rRNA* genotypes, but geography, climatic or edaphic factors did not influence the C-utilisation phenotypes. The C-utilisation groups were also not related to the pH-preference of the strains.

Some carbon sources (e.g. D-glucosamine, laminarin, L-pyroglutamic acid) were 'preferred' by the strains over others (e.g. citric acid, tyramine, oxalic acid). In general, most carboxylic acids (56%) did not stimulate growth over the controls, whereas most carbohydrates (89%) were utilised effectively. Four of the strains (4P1N1, 9P1N1, 46P2N1 and 74P3N2) used the most carbon sources and exhibited the highest growth (larger AUCs). Most strains (n = 16, 84%) were genetically similar to each other based on the *16S rRNA* and *nodC* sequencing results, i.e. no differences in their phylogenies.

4.4.2 *16S rRNA* and *nodC* phylogenies

Sequencing of the commonly used *16S rRNA* showed that, as expected, a majority (84%) of the strains were similar to each other and to *R. leguminosarum* type specimens in their *16S rRNA* genotypes. This result corresponded well with the preliminary sequencing results from Chapters 2 (NGS) and 3 (Sanger). The exceptions were strains 73P1N2 and 81P2N2 which did not cluster with the other 16 strains and shared a higher similarity with *Sinorhizobium* and *Agrobacterium*, respectively. Although strain 74P3N2 shared 99% sequence similarity with genus *Rhizobium*, it is likely that this strain is not a *leguminosarum* species, and hence it did not cluster with the other 13 strains. However, it was not unexpected to isolate these strains from nodules, since both *Agrobacterium* and *Sinorhizobium* have been previously recovered from nodules of clovers (Sturz et al. 1997; Wigley 2017).

The five *16S rRNA* genotypes (genetic distance) were not correlated with soil order, edaphic properties, sampling distance or climatic properties of sampling sites. This indicated that abiotic factors had not influenced the potentially ubiquitous *16S rRNA* genotypes found in pastoral sites around New Zealand. However, only a small number of strains were used for this comparison, thus a distinct association with abiotic factors may not become evident. The *nodC* sequences revealed a similar pattern to the *16S rRNA* sequences, i.e. the three strains from which *nodC* could not be amplified failed to nodulate white clover roots (data not shown). The taxa to which these strains belong have been isolated and/or identified from soils (Bashan and De-Bashan 2005; Wigley et al. 2017), thus it is likely for them to become passively entrapped during the nodule-

forming process. The other 16 strains showed *nodC* sequence similarity to *Rhizobium leguminosarum* bv. *trifolii*, which was also corroborated by the strains forming nodules on white clover roots (data not shown).

4.4.3 Carbon utilisation by strains

Although only 19 strains were profiled on their C-source utilisation, this is not dissimilar to the number of strains assessed in previous studies. The main point of difference between the present work and the previous studies on bacterial metabolism is that there was a deliberate selection of strains from a much larger group. The strains were selected from a diverse geographic and edaphic range, as well as from pre-screening of 138 strains on pH broths. Omnilog analysis identified patterns within the carbon utilisation that could be used to group strains. The amount of phenotypic variation was unexpectedly high among the 19 strains (10 groups were identified, Figure 4.4). Other studies have shown strains of *Rhizobium* spp. to demonstrate adaptation and/or tolerance to different physicochemical conditions, but relatively fewer metabolic groups were found amongst those strains. Rai *et al.* (2012) recovered 28 strains of *Mesorhizobium* spp. from chickpea root nodules collected from different geographic locations in India and phenotypically characterised them based on pH, temperature, and salt tolerance as well as phosphate solubilisation and antibiotic resistance. Their results showed that the 28 strains clustered into four distinct phenotypic groups (Rai et al. 2012). A study by Mazur *et al.* (2013) demonstrated metabolic versatility of 22 *R. leguminosarum* bv. *trifolii* strains (selected from 126 strains and which differed in their plasmid profiles) when they were subjected to phenotypic profiling (Biolog™ plates) comprising utilisation of C, N, P, and S sources as well as tolerance to pH and osmolytes. The 22 strains formed three major metabolic groups on the basis of substrate utilisation and/or tolerance to pH and osmolytes (Mazur et al. 2013).

This work has produced information on a large number of naturalised strains collected across a broad pH range in New Zealand. Some phenotypic variation was anticipated, but the rate of uniqueness was surprisingly high given that most strains were genotypically (*16S rRNA*) similar to each other. Furthermore, there was a strong association between the phenotypic variation and genetic distance of the five *16S rRNA* genotypes. It is likely that the ability of rhizobia to metabolise a variety of carbon compounds is based upon gene expression, i.e. up-regulation of existing chromosomal or plasmid genes that code for specific enzymes linked with carbon catabolism. This is supported by Black *et al.* (2012), who found a high number of (Kyoto Encyclopedia of Genes and Genomes; KEGG) protein orthologues in *R. leguminosarum* related to carbohydrate and amino acid metabolism pathways. Zahran (2017) suggested that phenotypic variation based on C-utilisation is likely related to plasmid-borne genes and not ones hosted in the core genome (*16S rRNA*), however, the results here suggest that the core genome may be involved in the phenotypic differences.

Strains demonstrated differences in the number and types of carbon sources they were able to metabolise, and there was a link between the number of different carbon sources utilised and increased growth for some strains. However, there was no association of C-utilisation phenotypes and pH-adaptation phenotypes (Table 4.5), which suggested other metabolic changes were responsible for pH adaptation in rhizobia. Graham (2008) stated that a wide range of environmental and edaphic factors may affect the phenotypic diversity of rhizobia species among soils. However, in the present study, there was no evidence of association of phenotypes with either soil order, sampling location and soil physicochemical or climatic properties. The carbon sources belonging to carbohydrate, alcohol and polymer groups produced the largest average AUCs among strains, whereas carbon sources which are amines, amino acids and esters produced, on average, the smallest AUCs (Table 4.3), these results are consistent with previous findings by Wielbo *et al.* (2010) and Mazur *et al.* (2013). However, the number of carbon sources in each of these groups are not the same and that may have influenced the statistics. Of the carbon sources that were main contributors towards the phenotype groupings (Table 4.6), hydroxy-L-proline, L-aspartic acid and L-pyroglutamic acid were ones which had greater contribution (> 10%) towards certain SIMPROF groups. These three amino acids provided a source of carbon, and are also important sources of nitrogen for the bacterial strains.

Differences in carbon utilisation between strains could be due to their ability to grow quickly (fast-growing strains) or slowly (slow-growing strains) on the substrates. Stowers (1985), in one of the seminal reviews on rhizobial carbon metabolism, mentioned that fast-growing rhizobia have the ability to utilise more carbon substrates than the slow-growing rhizobia. The four strains (4P1N1, 9P1N1, 46P2N1 and 74P3N2) which demonstrated high carbon usage as well as increased growth (Table 4.4) also had the highest AUC values when growing on m-inositol and L-rhamnose, compared with other strains. Two studies (reviewed in Prell and Poole, 2006) showed that mutants of *R. leguminosarum* bv. *viciae* and *R. leguminosarum* bv. *trifolii* deficient in catabolism of carbon sources such as inositol (Fry et al. 2001) or rhamnose (Oresnik et al. 1998) were less competitive for nodule formation. It was suggested that even small differences in the ability to use inositol or rhamnose may determine the strains of rhizobia which stay in the root-growth zone at the end of the infection thread and thus colonise the nodule tissue (Prell and Poole 2006). The work by Fry *et al.* (2001) and Oresnik *et al.* (1998) indicated that the genes coding for the catabolism of inositol and rhamnose are already present (and plasmid-borne) in *Rhizobium*. Thus, it is likely that the four strains possibly up-regulated the expression of those specific genes rather than acquiring new genes, and may have increased potential as commercial inoculants.

The carbon source D-glucosamine had the highest AUC among all strains, compared with all other carbon sources (Table A 24, Chapter 4 supplementary results). Nod factors consist of a backbone which contains N-acetyl-D-glucosamine (Lerouge et al. 1990), suggesting that D-glucosamine is an important carbon source in legume-rhizobia symbiosis. Dénarié *et al.* (1992) stated in their review that the *nodM* gene in *R. leguminosarum* bv. *trifolii* has homology to a D-glucosamine synthase.

One of those studies found that when *nodM* mutants of *S. meliloti* were exogenously supplied with D-glucosamine, it restored their ability to nodulate *Medicago* (Baev et al. 1991). Another study reported that N-acetyl-D-glucosamine was one of the most consumed carbon sources by bacterial communities derived from potato rhizospheres (İnceoğlu et al. 2012), which indicates that D-glucosamine could be a good additive to rhizobial seed-coat formulations.

The results suggested that mannitol, the widely used C-source for rhizobia, is not the most optimal for C utilisation by a wide number of strains. The carbon source D-mannitol stimulated, on average, lower growth in the strains compared with many other carbon sources, i.e. there were 48 (~40%) other carbon sources which had a higher average AUC across all strains than D-mannitol (Table A 23). This is an interesting finding, considering the most widely used semi-selective growth medium for rhizobia (Vincent 1970) contains mannitol as the sole carbon source. Perhaps, growth assays involving rhizobia could use carbon sources other than mannitol, which not only allow for better selection of *Rhizobium* spp., but also facilitate better growth. This may require further work to test if the growth-stimulating carbon sources are at least as effective as mannitol in selecting for rhizobia over other soil and/or nodule bacteria.

Only 19 strains were used for characterisation on the Omnilog system, thus this study warrants expansion. Strains recovered from nodules of subterranean clover (Chapter 3) could also be subjected to phenotypic profiling to assess if they demonstrate similar C-utilisation profiles to the WC strains. This study could also be expanded by including strains from more varied geographic sites, i.e. different (more) soil orders or physicochemical properties. The Omnilog system is robust and can generate big data relatively easily (Bochner 2008; Greetham 2014), but it is expensive and time consuming for many samples, especially when more replication is required. However, since this study has shown evidence for the Omnilog's efficacy in determining C-utilisation patterns, there is confidence that this work could be expanded to include more strains and more replication. This study was carried out iteratively to identify the effects of individual carbon sources on growth of rhizobia. However, this is not a good reflection of the nutritionally complex soil and/or plant root environment, where multiple carbon sources can influence rhizobial growth simultaneously (synergistically or antagonistically). Furthermore, not all the carbon sources present in the Biolog™ plates are necessarily found in the rhizosphere or root exudates, and so the effects of those carbon sources may end up being redundant in a field system. A study by Campbell *et al.* (1997) compiled a list of Biolog™ carbon sources that are also found in root exudates (probably from grassland soils in UK), which suggests that particular focus should be devoted to those carbon sources (as well as others found in the rhizosphere) rather than a broad range of carbon sources. However, if particular strains of rhizobia are able to effectively metabolise carbon sources that may not be normally found in the soil and/or plant root environment, then those carbon sources could be included in seed formulations with rhizobia strains to give them a competitive advantage over other bacteria.

4.5 Conclusions

This was a comprehensive study on the utilisation of 190 carbon sources by 19 rhizobia strains recovered from nodules of *T. repens* that had been selected based on evidence of pH-adaptation. This work showed:

- Based on *16S* and *nodC* phylogenies, 16 of the 19 strains tested were genetically similar.
- That the C-utilisation phenotypes of strains selected for their pH adaptation were highly diverse with the 19 strains being placed into 10 significantly different phenotypic groups, based on C-utilisation. Some strains also had increased carbon use efficiency and grew faster than others.
- Four of the strains (4P1N1, 9P1N1, 46P2N1 and 74P3N2) demonstrated higher fitness and flexibility in their utilisation of carbon sources compared to the other strains.
- That phenotypic variation was strongly associated with (*16S rRNA*) genetic distance between strains, but not geographic distance, edaphic and climatic properties.
- Despite a linkage between soil pH and C availability, there was no association between C-utilisation phenotypes and media-optimum pH of those strains.
- That some carbon sources (n = 69, 36%) did not support the growth of strains more than the negative (water) control, and these were predominantly amines and carboxylic acids.
- Support for previous research such that carbohydrates were the carbon sources used most commonly (n = 63, 33%).

5 Concluding Discussion

In recent years there has been increasing understanding of the importance of plant-microbe interactions in the ecological success of plants, and studies have characterised the microbial communities in plant tissues such as leaves and roots (Redford et al. 2010; Sun et al. 2010). The legume nodule is a specialised plant structure formed in response to interactions with nodulating bacteria. This structure has a central role in the transition of nitrogen from the atmosphere to the soil and thus, a key place in global nutrient cycles (Graham and Vance 2003). Emerging research has shown that in addition to rhizobia the nodule can contain other bacterial taxa, however, the significance of these communities and the factors that shape their composition is not well understood (papers cited in Martinez-Hidalgo and Hirsch (2017)). The overall aim of this PhD research programme was to address some of these gaps by determining the relationship between soil pH, a known primary driver of soil microbial communities, and the diversity of bacterial communities within the nodules of subterranean and white clovers. The key outcomes from this work were: i) that there is a relationship between soil pH and the community of bacteria in the nodules of SC and WC, ii) that pH adaptation of rhizobia inhabiting nodules was rare but correlated with the pH of the soil of origin and iii) that rhizobia demonstrating pH adaptation were diverse in their ability to utilise C-sources, but that the use of specific C-sources by the rhizobia did not correlate with the type of pH adaptation.

The research detailed in this thesis contributed to the MBIE programme “Improved forage legume-rhizobia performance”. This MBIE programme addresses the discrepancy between the deliberate selective breeding of clover cultivars and the very limited selection of commercial rhizobial inoculants. The current approach in New Zealand for commercial inoculation of clover is the application of strains TA1 for white clover (WC) and WSM1325 for subterranean clover (SC), onto the surface of seed, with strain TA1 having been used since the 1960’s. As a result, single strains of rhizobia of a standard genetic and physiological potential have been introduced into the diverse range of soil habitats present across New Zealand. Similar to soils around the world, New Zealand pastures vary widely in their properties, spanning a range of basic chemical and fertility conditions. It is, therefore, impossible that a single *Rhizobium* strain would be ideally matched to all pasture conditions. Paradoxically, this is not expected of clover or other forage plants, where cultivars are specifically bred for different soil and environmental conditions. In contrast, there are few parallel rhizobial germplasm assessments. The soil into which these strains have been introduced is a “natural laboratory”, placing a selective pressure on the rhizobia strains. Assessing the relationship between the properties of natural soils, the phenotypes of rhizobia and their nodulation success could therefore provide new understanding of rhizobial ecology in the complex soil background. In this research, soil pH was chosen as the main edaphic factor of interest due to its clear role in shaping soil microbial communities (Wakelin et al. 2008; Lauber et al. 2009; Rousk et al. 2010) and its significance for nodulation and plant growth. In New Zealand, low (acidic) pH soils are a particular problem for rhizobia in productive settings, negatively

affecting clover-rhizobia symbiosis and thereby clover growth. Thus, an applied outcome of this project was to demonstrate the potential for selection of specialised strains for use in soils with marginal pH levels (e.g. pH <5.5).

The main outcome of Chapter 2 was to demonstrate that there was a relationship between the bacterial community present in the nodule and the pH of the soil of origin. This showed, for the first time, that plant-microbe interactions within the nodule are influenced by soil pH and suggested a functional significance. Although the host, rather than soil pH, was the strongest driver of the nodule community, with SC nodules exhibiting a higher alpha diversity than WC nodules. This thesis describes the first use of amplicon sequencing (Illumina HiSeq) to study the bacterial community inside the nodules of SC and WC when grown in soils spanning a broad pH range (Chapter 2). The results confirmed previous studies (Marilley and Aragno 1999; Liu et al. 2007; Hartman et al. 2017) that although *Rhizobium*, *Bradyrhizobium* and *Sinorhizobium* genera were the most dominant taxa identified in nodules, non-*Rhizobium* genera were also present. The non-*Rhizobium* genera included those previously identified within clover nodules (Hartman et al. 2017; Ramana 2018), among others. To determine whether these other genera have beneficial properties, functional studies could be performed to identify their roles *in planta*. In the current study no genera were completely excluded from either clover species, with the main difference being the relative abundance of those bacteria. If the non-*Rhizobium* bacteria were actively recruited it would suggest that they may in fact benefit the clover in some way, but studying the underlying mechanism(s) is difficult. Such work could determine if there is a synergistic relationship (Martínez-Hidalgo and Hirsch 2017) between the 'other' genera and *Rhizobium* inside clover nodules and, if so, what minimum numbers of these bacteria are required to initiate a positive association with *Rhizobium*.

The work presented in Chapter 2 also demonstrated that the two clover species were different in their association with bacteria in their nodules, with the relative abundance of non-*Rhizobium* bacteria higher in SC nodules than in WC nodules, irrespective of soil pH. Thus, there was evidence to show that the host species was a bigger driver influencing bacterial communities than soil pH. Higher alpha diversity of total nodule microbiome was found in nodules of SC grown in soils of pH group A compared with those grown in either groups B or C soils. This raised two possibilities: i) there is more damage or crack formation in roots in pH-stressed soils, which may allow the wider soil microbiome to access the nodules; or more likely ii) that there is active recruitment of bacteria by SC plants growing in pH-stressed soils. The latter explanation is supported by observed differences between the two host species, despite both experiencing the same soil pH stress, and by the work of Seth (2017), on selective recruitment of phosphate-solubilising rhizobia into clover nodules, that showed the nodule community does not simply reflect the rhizosphere community. If there is a link between the presence of non-*Rhizobium* bacteria in SC nodules and the ability of SC plants to survive in low pH soils, then there is potential to improve the persistence of SC in acid soils by augmenting with bacteria that help the plants

cope in pastoral land affected by soil acidity. It is important to note that only one cultivar of each clover was used for the experiments ('Denmark' for SC and 'Tribute' for WC), thus these results may not represent overall differences between the species because of cultivar effects. Greater depth on the potential for SC as a clover species to recruit useful species of non-rhizobia (e.g. *Pseudomonas*) to nodules could be generated by examining a range of cultivars under acidic soil conditions. Future work assessing the metagenome or transcriptome (reviewed in Barret *et al.* (2011)) in nodules in response to pH stress may provide insights into the mechanisms by which genera other than *Rhizobium* provide benefits to the clover.

This work produced new information on the association of *nodC* diversity in clover nodules with the pH of the soil. Despite sequencing only a small portion of the *nodC* gene (corresponding to 113 amino acids, ~9% of WSM1325 NodC), a high diversity (n = 353) of NodC protein sequences was found, and this has not been previously reported. Of the representative sequences, some *nodC* genotypes associated strongly with pH, others associated strongly with host and a few were found at higher frequencies in a particular host at a specific soil pH. The similarity between the 353 NodC protein sequences and their placement into two major groups, with the commercial strains TA1 and WSM1325 each placed in one of the groups, suggested that these commercial strains are the progenitors of the nodule inhabitants. These findings and those of Zhang et al. (2016) may allow the inference to be drawn "that horizontal gene transfer between closely related species possibly directed the diversification and evolution of clover-nodulating rhizobia to help the clover plants colonise diverse environments". Further work to assess the diversity of nodulation proteins within the soil environment and their association with edaphic factors could assess other important symbiotic genes, such as *nodA* and *nodD* (QQAD, Young *et al.*, unpublished) or additional regions of *nodC*. Greater analysis could identify variants that have a preference for a particular host/pH combination and could be extended to explore the influence of cultivar. Ultimately, use of these genetic signatures may provide a more rapid, accurate or cost-effective process for selection of pH- or host-adapted strains.

Chapter 2 demonstrated the presence of high rhizobia diversity in nodules, with some genotypes present in different proportions according to host and soil pH. This raised the possibility that soil pH drives adaptation in the free living rhizobia community to sub-optimal pH. Therefore, Chapter 3 assessed the relationship between pH of soil of origin and rhizobia pH preference. To achieve this, a novel *in vitro* bioassay was developed for fast and easy screening of growth of a large number of strains (n = 299) cultured from nodules of SC and WC. One of the major outcomes of Chapter 3 was that pH adaptation was rare in rhizobia, with only few strains adapted to pH < 5.5 and pH > 7.5 were found (n = 4, SC; n = 3, WC), indicating that pH adaptation is either not strongly selected for or is difficult to achieve for rhizobia. There was a correlation between the media-optimum pH and the pH of the soil or origin for strains of both clover species. However, this correlation seemed to be driven by strains which originated from soils of pH 7.1 and 7.5, which hints more towards alkaline-adaptation than acid-adaptation. This type of laboratory screening

test was a quick and easy way to identify strains with potential to grow at sub-optimal pH and could be applied to identify new commercial strains for marginal pastoral sites (Dilworth et al. 2001). Unsurprisingly, a majority of the strains (~88%) showed optimal growth in liquid broths of pH 5.8–7.5, since this range is most suitable for the growth of bacteria. This finding fitted Connell's (1978) intermediate disturbance hypothesis that "diversity is higher when disturbances are intermediate on the scales of frequency and intensity". The hypothesis discusses diversity, disturbance and intensity; these can be respectively related to population of pH-adapted strains, pH-stress and level of acidity or alkalinity.

The results of Chapter 3 identified strains that may have potential as new commercial inoculants for achieving more optimal symbiotic outcomes with clover in acidic soil pH. Although beyond the scope of the ecology focus of this thesis, this work could progress in a similar way to the work of Ballard *et al.* (2003), who found that naturalised acid tolerant *S. meliloti* strains formed effective symbioses with lucerne, in terms of shoot dry matter and nodule formation. They showed that when combined with an acid tolerant lucerne cultivar, significant gains in establishment and growth in low pH soils were possible. In a similar way, further assessments could be conducted to identify pH-adapted genotypes within each clover species/cultivar. To further explore their efficacy in a soil environment the selected strains from SC and WC could be tested in a soil system where the biological background has been removed, such as via gamma irradiation treatment, to assess nodulation (size, number) and N-fixation (clover biomass, leaf greenness) potential, as well as competition for nodule initiation against commercial/other strains. To be considered commercially useful the selected strains would have to stimulate clover growth equally to TA1 or WSM1325, but they will have the added advantage of being able to tolerate more stressful soil environments than the commercial inoculants.

There is a known relationship between availability of carbon sources and soil pH (Kemmitt et al. 2006; Rousk et al. 2009). Prior published works have explored the ability of strains of rhizobia to metabolise C-sources using Biolog™ plates (Wielbo et al. 2010; Mazur et al. 2013). However, these studies predominantly assessed rhizobia strains that were reflective of the majority. The work presented here is the first to specifically focus on rare strains with demonstrated pH adaptation. The work showed that there was a high degree of diversity amongst the pH-adapted strains, with 10 phenotypic (metabolic) groups from 19 strains. This new understanding of the degree of diversification in C-utilisation provides a fundamental basis for formulation technologies. Information from growth assays of strains on a variety of carbon sources from different groups (such as amino acids, carbohydrates, carboxylic acids, fatty acids and alcohols) may help in delivery of selected strains into soils. Using knowledge of carbon-source preferences, strain-specific formulations could be developed to incite maximum growth, which may improve successful delivery of strains to the field. Carbon substrates which can specifically stimulate the growth of a selected commercial strain may provide a competitive edge when applied to a soil system. For example, strain 79P2N2 with pectin, strain 9P1N1 with D-ribose or strain 1P1N1 with

acetoacetic acid. This is of importance in New Zealand due to the large range of naturalised strains which are present in soils at populations between 10^3 and 10^8 per gram of soil (Rys and Bonish 1981; Wakelin et al. 2018). Prior work has shown that the presence of these large naturalised communities is a major impediment to the success of commercial inoculants (Wakelin et al. 2018).

Chapter 4 demonstrated that there was variation in the way in which strains isolated from WC nodules utilise carbon sources. Strains which utilise a wider range of carbon sources, or that are able to demonstrate higher carbon use efficiency, might be more competitive than their counterparts in soil (Wielbo et al. 2007). If the same strains also demonstrated an ability to promote plant growth (greater symbiotic potential), then their usefulness as effective commercial inoculants would increase. Knowledge on carbon metabolism by rhizobia strains may enable the pH bioassay (Chapter 3) to be improved by replacing YMB (mannitol) with medium that is not only semi-selective for rhizobia, but also improves growth of strains (e.g. by including glucosamine or laminarin). The carbon utilisation profiles of selected strains (from the pH bioassay) isolated from SC nodules should also be assessed, and analysed for associations between phenotypes, (*16S rRNA*) genetic distance, soil and climatic properties. These could provide a new source of bio-inoculants for challenging soil environments. Information on plasmid profiles (using Eckhardt gels or sequencing of the plasmids) of selected strains will be useful to elucidate if pH-adaptation and variation in C-utilisation is linked to the ancillary and/or core genome. The genome of several *R. leguminosarum* strains with known phenotypes have been sequenced (Young et al. 2006). These provide a template for comparison of metabolic pathways among strains, potentially implicated in adaptation to soil pH by genetic variation or the presence or absence of chromosomal and plasmid genes. If identified, the DNA sequence, regulatory elements and expression profiles of key genes in these pathways could be examined under different treatments, i.e. pH levels, geography, climate.

The phenotyping undertaken in this research was limited to carbon usage, because of the association between soil pH and carbon availability and the diversity of C-sources in the different niches occupied by rhizobia. In future work, additional phenotyping and gene functionality assessments could be carried out on the cohort of strains demonstrating commercialisation potential. GeoChip is a functional gene microarray technology and a powerful high-throughput tool for studying microbial community functional structure, and linking microbial communities to ecosystem processes and functioning (Zhang et al. 2013). GeoChip 3.0 has been developed with ~28,000 probes covering approximately 57,000 gene variants from 292 functional gene families involved in C, N, P and S cycles, energy metabolism, antibiotic resistance and metal resistance (He et al. 2010). This could provide genetic information about selected strains linking to their pH-adaptation, C-utilisation and symbiotic potential.

In summary, this study has provided new knowledge about rhizobial ecology with respect to soil pH as an important environmental gradient. It adds to growing understanding of the importance of plant-microbe interactions and addresses a knowledge gap for nodules as specialised and important plant structures. Comparisons between the bacterial communities dwelling in nodules of two important clover species were made, with new data generated on *nodC* diversity and *16S rRNA* relative abundance. There was evidence to support a relationship between soil pH and the nodule microbiome, with clover species playing a dominant role in shaping the nodule bacterial community. The research demonstrated that pH adaptation by rhizobia was rare, but correlated with pH of soil of origin. A cohort of strains which demonstrated adaptation to pH <5.5, pH >7.5 and those adapted to a broad pH range (4.5 – 9.0) were identified. Due to the relationship between soil C availability and soil pH, selected strains from WC nodules were further screened on their ability to metabolise various carbon sources. A larger amount of phenotypic diversity was observed than had previously been described for rhizobia, which was correlated with genetic distance only and not soil or climatic factors. Overall, this new information will be crucial in designing other downstream experiments which will eventually help to improve the delivery and establishment of new and effective rhizobial inoculants in New Zealand pastoral soils.

Appendix 1: Media and reagents

Table A 1: Yeast Mannitol Agar (YMA)

Chemical	**Amount**
Yeast extract	1 g
Mannitol	4 g
Dipotassium hydrogen orthophosphate (K_2HPO_4)	0.5 g
Magnesium sulphate ($MgSO_4$)	0.2 g
Sodium chloride (NaCl)	0.1 g
Bacteriological agar	15 g
H_2O	1 L

- Autoclave for 15 min at 121°C and 15 KPa

Yeast Mannitol Broth (YMB) was made using the same ingredients as Table A 1, except for bacteriological agar. It was autoclaved for 15 min at 121°C and 15 KPa

Table A 2: McKnight's Nutrient Solution

Chemical	**Amount**
Calcium chloride ($CaCl_2$)	0.25 ml
Magnesium sulphate ($MgSO_4$)	0.25 ml
Potassium dihydrogen orthophosphate (KH_2PO_4)	1 ml
Potassium chloride (KCl)	1 ml
Trace elements stock solution *	0.5 ml
EDTA – $FeCl_3$ stock solution **	0.75 ml
Sodium hydroxide (NaOH)	0.25 ml
Millipore H_2O	1000 ml

* Trace elements stock (1 L) made up of:

Boric acid (H_3BO_3)	2.86 g
Manganese sulphate ($MnSO_4.4H_2O$)	2.03 g
Zinc sulphate ($ZnSO_4.7H_2O$)	0.222 g
Copper sulphate ($CuSO_4.5H_2O$)	0.079 g
Molybdic acid ($H_2MoO_4.H_2O$)	0.09 g

** EDTA – $FeCl_3$ stock solution (1 L) made up of:

Ethylene diamine tetra-acetic, sodium salt (EDTA)	2 g
60% w/v Ferric chloride solution ($FeCl_3$)	16.8 mL

- Autoclave mixed solution for 15 min at 121°C and 15 KPa

TAE buffer

A 50× stock solution was prepared by combining 242 g Tris base in water, adding 57.1 mL glacial acetic acid, and 100 mL of 500 mM EDTA (pH 8.0) solution. The final volume was brought up to 1 L. This stock solution was diluted 49:1 with RO H_2O to make a 1× working solution.

Appendix 2: Supplementary sequencing data

PCR Primers (5' to 3')

***16S rRNA* (Weisburg et al. 1991)**

F27: AGA GTT TGA TCM TGG CTC AG

R1494: CTA CGG YTA CCT TGT TAC GAC

***nodC* (Laguerre et al. 2001)**

nodCF: AYG THG TYG AYG ACG GTT C

nodCR: CGY GAC AGC CAN TCK CTA TTG

A, C, G, T = standard nucleotides; M = C or A; Y = C or T; R = A or G; S = G or C; B = T or C or G; H = A or C or T; N = A or G or C or T; K = T or G

PCR protocols

Table A 3: *16S rRNA*

No. of cycles	Annealing temperature	Time
1 ×	94˚C	3 min
35 ×	94˚C	30 s
	55˚C	30 s
	72˚C	1 min
1 ×	72˚C	7 min
1 ×	4˚C	∞

Table A 4: *nodC*

No. of cycles	Annealing temperature	Time
1 ×	95˚C	3 min
35 ×	94˚C	30 s
	55˚C	30 s
	72˚C	45 s
1 ×	72˚C	7 min
1 ×	4˚C	∞

NGS information

Table A 5: List of HPLC-purified *16S rRNA* primers used for the NGS experiment.

Primer name (barcode + direction)	Primers (5'- 3') (Barcode Linker Primer)
2 + F341	CGATGT GA CCTAYGGGRBGCASCAG
5 + F341	ACAGTG GA CCTAYGGGRBGCASCAG
17 + F341	GTAGAG GA CCTAYGGGRBGCASCAG
36 + F341	CCAACA GA CCTAYGGGRBGCASCAG
38 + F341	CTAGCT GA CCTAYGGGRBGCASCAG
44 + F341	TATAAT GA CCTAYGGGRBGCASCAG
1 + R806	ATCACG GT GGACTACNNGGGTATCTAAT
3 + R806	TTAGGC GT GGACTACNNGGGTATCTAAT
11 + R806	GGCTAC GT GGACTACNNGGGTATCTAAT
14 + R806	AGTTCC GT GGACTACNNGGGTATCTAAT
26 + R806	ATGAGC GT GGACTACNNGGGTATCTAAT
37 + R806	CGGAAT GT GGACTACNNGGGTATCTAAT

Table A 6: List of HPLC-purified *nodC* primers used for the NGS experiment.

Primer name (barcode + direction)	Primers (5'- 3') (Barcode Linker Primer)
1 + F10	ATCACG GA CTACGCCGCATCGTATCAGT
17 + F10	GTAGAG GA CTACGCCGCATCGTATCAGT
19 + F10	GTGAAA GA CTACGCCGCATCGTATCAGT
2 + F10	CGATGT GA CTACGCCGCATCGTATCAGT
36 + F10	CCAACA GA CTACGCCGCATCGTATCAGT
38 + F10	CTAGCT GA CTACGCCGCATCGTATCAGT
4+ F10	TGACCA GA CTACGCCGCATCGTATCAGT
5 + F10	ACAGTG GA CTACGCCGCATCGTATCAGT
11 + R357	GGCTAC GT GAGGACCGTCATCTGACCAT
14 + R357	AGTTCC GT GAGGACCGTCATCTGACCAT
26 + R357	ATGAGC GT GAGGACCGTCATCTGACCAT
37 + R357	CGGAAT GT GAGGACCGTCATCTGACCAT
6 + R357	GCCAAT GT GAGGACCGTCATCTGACCAT
7 + R357	CAGATC GT GAGGACCGTCATCTGACCAT
8 + R357	ACTTGA GT GAGGACCGTCATCTGACCAT

Table A 7: QC report for DNA quality from Novogene

Novogene ID	Sample Name	Conc. (ng/μl)	Volume (μL)	Amount (μg)	Conclusion
TD171108840	16S 1	22.60	203	4.590	Pass
TD171108841	16S 2	23.76	120	2.850	Pass
TD171108842	16S 3	31.20	132	4.120	Pass
TD171108843	16S 4	27.60	64	1.770	Pass
TD171108844	16S 5	27.40	62	1.700	Pass
TD171108845	16S 6	19.60	120	2.350	Pass
TD171108846	16S 7	17.50	188	3.290	Pass
TD171108847	16S 8	17.80	132	2.350	Pass
TD171108848	nodC 1	17.40	142	2.470	Pass
TD171108849	nodC 2	23.80	114	2.710	Pass
TD171108850	nodC 3	19.10	119	2.270	Pass
TD171108851	nodC 4	21.34	98	2.090	Pass
TD171108852	nodC 5	24.00	86	2.060	Pass
TD171108853	nodC 6	19.50	101	1.970	Pass
TD171108854	nodC 7	13.90	104	1.450	Pass
TD171108855	nodC 8	15.90	99	1.570	Pass

nodC database (62 sequences)

>D28960.1 *Rhizobium leguminosarum* bv. *viciae* nodC gene, partial cds, strain:USDA2478
CAAAAATCGCGACGCGGTGGAGGCTCAGCGCGCTGCCTATGCAGACGATGAGAGATTCAAATTCACAATT
CTCCCTAAAAATGTTGGAAAGCGCAAAGCGCAAATCGCCGCTATAACCCAGTCCTCTGGGGACCTCATCT
TGAATGTGGACTCAGACACCACGATCGCCCCCGACGTCGTCTCGAAGCTTGCCCATAAAATGCGCGATCC
AGCAGTCGGTGCGGCGATGGGCCAAATGAAAGCCAGTAACCAGGCGGACACCTGGCTAACTCGC

>D28959.1 *Rhizobium leguminosarum* bv. *trifolii* nodC gene, partial cds, strain:USDA2161
CAGAAATCGCGACGCCGTCGTGGCCGAGCAGCTTGCCTATGCGGGCGACGCGAGATTCGAATTCATCATG
CTTCCCAGGAATGTCGGAAAGCGCAAAGCGCAAATCGCTGCCATTTCCCGATCATCCGGGGACCTGATCC
TGAACGTGGATTCAGACACCACTCTCGCCTCCGACGTCGTTTCAAAACTGAGTCAGAAAATGCGCGATCC
GGCGGTCGGTGCGGTGATGGGGCAGCTTGTAGCGAGCAATCAGAGCGACTCGTGGCTGACCCGA

>D28957.1 *Bradyrhizobium japonicum* nodC gene, partial cds, strain:USDA136
GGCAAACCGCGACGTTGTCGCGCTTGTACACCGGATATATGCGAGCGATCCGAGATTCAGTTTTATCTTG
TTGGCGAACAATGTGGGAAAGCGCAAGGCGCAGATCGCAGCGATACGCAGCTCATCCGGTGATCTGGTTC
TCAACGTCGATTCCGATACGATACTTGCTGCCGACGTCGTCACGAAGCTTGTATTGAAGATGCATGACCC
GGGAATCGGTGCGGCCATGGGTCAGCTGATCGCGAGCAATCGCAACCAGACCTGGCTGACCAGG

>D28961.1 *Rhizobium leguminosarum* bv. *phaseoli* nodC gene, partial cds, strain:USDA2676
CGAAAATCGCGAGGCTTTGCAACTTGTGCACGAGGCCTTCGCACGAGACCCCAGATTCAATATTCTCGTG
CTTCCCCAGAATGTCGGCAAACGGAAGGCACAGATCGCTGCGATACGCCGCTCTGCTGGAGATATGGTGT
TAAACGTCGACTCCGACACAATCCTCGCATCTGATGTCATCAGGAAGCTCGTGCCTAAAATGCAAGATGC

GGCTGTCGGCGCGGCCATGGGACAGTTGACGGCCCGCAACCGAAGCGACAGTTGGCTGACCCGT

>AY665788.1 *Rhizobium leguminosarum* isolate PS25-2 NodC (nodC) gene, partial cds
GCCTGCAACGAGGAGCGCGCGGCACAAGCTCGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCATGTGCGA
TGTACCGTCGGTCTGCTATGCTTTCGCTGCTCGATCAGTACGAGACGCAGCTTTATCGCGGCAAGCCGAG
TGACTTCGGCGAAGATCGCCATTTGACGATTCTCATGCTGAGCGCAGGCTTTCGAACTGAGTATGTTCCG
AGTGCCATCGCGGCGACAGTCGTTCCAGACACGATGGGTGTTTATCTACGCCAACAACTACGGTGGGCAC
GCAGCACCTTTCGGGATACTTTGCTTGCGCTTCCTGTACTGCCTGGTCTCGATCGGTATCTCACGCTGGA
CGCAATAGGGCAAAATGTCGGCCTGCTACTTCTTGCGCTGTCGGTATTGACAGGAATTGGCCAGTTTGCG
CTGACCGCCACACTGCCCTGGTGGACGATCCTGGTCATTGGATCCATGACTCTTGTACGCTGTAGCGTGG
CTGCCTATCGCGCCCGCGAACTTAGGTTTTTGGGTTTTGCTCTCCACACGCTCGTGAACATCTTTCTCTT
AATTCCCTTGAAGGC

>AY664625.1 *Rhizobium* sp. VR14-1 NodC (nodC) gene, partial cds
CTTCGGGGAAGACCGCCACCTCACAATCCTCATGCTAAACGCAGGCTTTCGAACCGAGTACGTTCCGGAA
GCCATCGCGGCGACCGTCGTTCCAAACACGATGGCGGCGTATCTGCGGCAACAACTACGCTGGGCACGCA
GCACGTTTCGTGACACATTGCTCGCGCTGCGCCTACTGCCGGGCCTCGATCGCTATCTTACGCTGGACGT
GATCGGACAGAATCTTGGTCCGCTGCTCCTCGCCATCTCGGTGCTAACGGGGCTAGCACAGCTCGCTCTA
ACAGCCACAGTGCCCTGGTCGACGATCCTGATGATTGCATCTATGACAATGGTCCGCTGCGGCGTAGCGG
CCTTTCGAGCGCGCCAGCTGCGTTTCCTTGGGTTTTCGCTTCACACCCTCGTCAACATCACTCTCCTGCT
CCCCCTC

>AY664622.1 *Rhizobium leguminosarum* isolate VF25-1 NodC (nodC) gene, partial cds
CTGCAACGAGGAGCGCGCGGCACAAGCTCGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCATGTGCGATG
TACCGTCGGTCTGCTATGCTTTCGCTGCTCGATCAGTACGAGACGCAGCTTTATCGCGGCAAGCCGAGTG
ACTTCGGTGAAGATCGCCATTTGACGATTCTCATGCTGAGCGCCGGCTTTCGAACTGAGTATGTTCCGAC
CGCCATCGCGGCGACAGTCGTTCCAGACACAATGGGTGTTTATCTGCGTCAACAACTCCGGTGGGCACGC
AGCACCTTTCGGGATACTTTGCTTGCGCTCCCCATACTGCCTGGCCTCGATCGGTATCTCACTCTGGACG
TAATCGGGCAAAATGGCGGCCTTCTGCTTCTTGCGCTGTCGGTATTGACGGGTATTGGCCAGTTTGCGCT
GACCGCCACAGTACCATGGTGGACGATCCTGGTGATCGGATCCATGACCCTTGTACGGTGTAGCGTAGTT
GCCTATCGCGCCCGCGAACTTAGGTTTCTGGGTTTTGCTCTCCACACGCTCGTGAACATCTTTCTCTTAA
TCCCT

>AY664618.1 *Sinorhizobium meliloti* isolate MS25-3 NodC (nodC) gene, partial cds
GCGGCCCTTGTGCTATGTACCGTCGGTCGGCGCTCGCTTCGCTGCTTGACCAGTACGAAACGCAACTGTT
TCGCGGTAAGCCAAGCGACTTCGGTGAGGACCGCCATCTGACGATTCTCATGTTGAAGGCAGGCTTTCGA
ACTGAGTACGTTCCAGACGCCATAGTGGCAACCGTCGTCCCGGATACGCTGAAACCATATCTGCGCCAAC
AACTGCGTTGGGCACGCAGCACGTTCCGTGACACGTTTCTAGCGCTCCCTCTGTTGCGCGGCCTCAGCCC
TTTTCTCGCATTTGACGCGGTCGGACAGAATATCGGGCAACTGTTGCTCGCCCTGTCGGTGGTGACGGGT
CTTGCGCATCTCATAATGACCGCCACAGTGCCATGGTGGACAATTTTGATTATTGCGTGCATGACCATTA
TACGCTGCAGCGTCGTAGCATGGCATGCTCGCCAACTTAGATTTCTTGGCTTCGTTCTGCACACACCCAT
CAACCTCTTTCTCATACTTCCGCTGAAAGCTT

>AY664610.1 *Sinorhizobium meliloti* isolate MS14-2 NodC (nodC) gene, partial cds
CTTCGGTGCTGTTATGTGTTGCTGCGGCCCTTGTGCTATGTACCGACGGTCAGCGCTCGCTTCGCTGCTT
GACCAGTACGAAACGCAACTGTTTCGCGGTAAGCTAAGCGACTTCGGTGAGGACCGCCATCTGACGATCC

TCATGTTGAAGGCGGGGTTTCGAACTGAGTATGTTCCAAACGCCATAGTGGCAACCGTTGTCCCGGATAC
GCTGAAATCTTATCTGCGCCAACAACTGCGTTGGGCACGCAGCACGTTCCGGGACACATTTCTAGCGCTC
CCTCTGTTGCGCGGCCTCAACCCTTTTCTCACATTTGACGTGGTCGGGCAGAATATCGGGCCGCTGTTGC
TTGCTCTGTCGGTGGTGACGGGACTTGCGCATTTCATAATGACCGCCACAGTGCCATGGTGGACAATTTT
GATTATTGCGTCCATGACCATTATACGCTGCAGCGTCGTAGCATTGCATGCTCGCCAACTTAGATTTCTT
GGCTTCGTTCTTCACACACCCATCAACCTCTTTCTTATACTTCCGTTGAAAGCTTATGCG

>AY664606.1 *Mesorhizobium* sp. CA14-1 NodC (nodC) gene, partial cds
TTGGTGCGGTCATGTGCTGCTGCGGCCCATGTGCAATGTACCGCCGGTCTTGTCTCCTTTCTCTGCTGGA
TCAATACGAGACGCAGCTGTTCAGGGGGAAACCAAGCGACTTCGGTGAAGACCGTCATCTTACGATCCTC
ATGCTGAAAGCAGGCTTTCGAACCGAGTACGTTCCAGGTGCCGTCGCGGCGACAGTCGTTCCGGACAAGA
TGGGACCTTATTTGCGCCAACAACTCCGCTGGGCACGGAGCACTTTCCGGGACACGATGCTTGCGCGAGG
TCTACTGCGTGGCCTGGATCGCTATCTAACTTTGGATGTGATGGGAGAGAATCTCGGCCCATTGTTGCTC
GGCATCGCGGTAGTAACGGCGCTCGGTGAGCTACTATTTTCACATACAGTAAATTGGTGGACGGTGCTGG
CTATTGTCACCATGACCATAATCAGATCTACGGTGTTATCTTTCCGCACTCGGCAGCTTCGATTCATCGG
CTATTCAATGCACGCGTCAATCAGGATATTCCTGTTGATCCC

>AY664604.1 *Bradyrhizobium* sp. VR25-1 NodC (nodC) gene, partial cds
CGCGCGGCACAGGCGCGCTTCGGTGCCGTCATGTGTTGCTGCGGCCCATGTGCAATGTATCGGCGTTCCG
CACTCACCTTGCTTCTTGATCAATACGAAGCGCAATTCTTTCGTGGGAAGCCGAGTGATTTCGGCGAGGA
CCGCCATCTAACGATACTCATGCTCAAGGCGGGGTTTCGAACCGAATACGTCCCAGACGCGATAGCAGCC
ACCGTCGTCCCGCACAGTCTTGGGCCATATCTGCGTCAGCAGCTCCGTTGGGCGCGCAGTACCTTTCGAG
ATACGTTTCTTGCATTGCGCCTGCTGCCAGAGCTCGATGGTTATTTGACCCTAGATGTTATCGGGCAAAA
TCTCGGCCCATTTCTCCTTGCCGTTTCAACACTTGCTGCGCTTGCACAGCTCGTGATTGGCGGCTCTATA
CCCTGGTGGACGGGACTGACGATTGCTGCAATGACTATGGTCCGGTGCAGTGTGGCAGCAATTCGTGCTC
GCGATTTGCGGTTTATCGGCTTCTCGCTCCATACACCGATCAATATCTTTCTATTACTACCATTGAAGGC
CTATGCGC

>AY664623.1 *Rhizobium leguminosarum* isolate VF25-2 NodC (nodC) gene, partial cds
CTGCACGAGGAGCGCGCGGCACAAGCTCGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCATGTGCGATGT
ACCGTCGGTCTGCTATGCTTTCGCTGCTCGATCAGTACGAGACGCAGCTTTATCGCGGCAAGCCGAGTGA
CTTCGGTGAAGATCGCCATTTGACGATTCTCATGCTGAGCGCCGGCTTTCGAACTGAGTATGTTCCGACC
GCCATCGCGGCGACAGTCGTTCCAGACACAATGGGTGTTTATCTGCGTCAACAACTCCGGTGGGCACGCA
GCACCTTTCGGGATACTTTGCTTGCGCTCCCCATACTGCCTGGCCTCGATCGGTATCTCACTCTGGACGT
AATCGGGCAAAATGGCGGCCTTCTGCTTCTTGCGCTGTCGGTATTGACGGGTATTGGCCAGTTTGCGCTG
ACCGCCACAGTGCCATGGTGGACGATCCTGGTGATCGGATCCATGACCCTTGTACGGTGTAGCGTAGTTG
CCTATCGCGCCCGCGAACTTAGGTTTCTGGGTTTTGCTCTCCACACGCTCGTGAACATCTTTCTCTTAAT
TACCCTGAAGGCCTATGCCCTT

>EF067926.1 *Bradyrhizobium* sp. CCBAU 33037 NodC (nodC) gene, partial cds
TGCAACGAAGAGCGTGCGGCACAGGCTCGTTTCGGTGCCGTCATGAGCTGCTGCGGCCCATGTGCGATGT
ATCGTCGCTCCGCGCTCGCTTTGCTGCTTGATCAATACGAAGCGCAATTCTTTCGCGGGAAGCCGAGCGA
TTTCGGCGAGGATCGCCATCTAACGATCCTCATGCTCAAGGCGGGGTTTCGAACCGAATATGTCCCGGAC
GCGATCGCAGCAACAGTAGTTCCGGACAGCCTTGGGCCATATCTGCGTCAGCAGCTCCGCTGGGCGCGCA
GTACCTTTCGGGACACGTTTCTTGCATTACGTCTGCTGCCAGAGCTCGATGGCTATCTGACGCTAGACGT
TATCGGGCAAAATCTCGGCCCGTTGCTTCTCGCCCTTTCATCACTGGCTGCGCTCGCGCAGCTTGTAATC

GGCGGCTCGGTACCCTGGTGGACGGGATTGACGATTGTAGCAATGGCGATGGTCCGGTGCAGTGTTGCAG
CGCTTCGTGCTCGCGAAGTGCGGTTTCTCGGCTTCTCGCTCCACACACCGATCAATATCTTTCTCTTACT
TCCTTTGAAGGCCTATGCACTTTGTACATTGAGCAATAGCGACTGGCTGTCGCG

>KP754750.1 *Rhizobium leguminosarum* bv. *viciae* strain WBAV_nodC_35, partial cds
CGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCATGTGCGATGTACCGTCGGTCTGCTATGCTTTCGCTGC
TCGATCAGTACGAGACGCAGCTTTATCGCGGCAAGCCGAGTGACTTCGGTGAAGATCGCCATTTGACGAT
TCTCATGCTGAGCGCAGGCTTTCGAACTGAGTATGTTCCGAGCGCCATCGCGGCGACAGTCGTTCCAGAC
ACAATGGGAGTTTATCTGCGTCAACAACTACGGTGGGCACGCAGCACTTTTCGGGATACATTGCTTGTGC
TTCCCGTACTACCTGGTCTCGATCGGTATCTCACGCTGGACGCAATCGGGCAAAATGTCGGCCTTCTACT
TCTTGCGCTGTCGGTATTGACAGGAATTGGTCAGTTTGCGCTGACCGCCACAGTACCCTGGTGGACGATC
TTGGTCATTGGATCCATGACTCTTGTACGGTGTAGCGTGGCTGCCTATCGCGCCCGCGAACTTAGGTTTT
TGGGTTTTGCTCTCCACACGCTCGTGAACATCTTTCTCTTAATTCCCTTGAAGGCCTATGCCCTTTGTAC
CTTGTC

>KP754742.1 *Rhizobium leguminosarum* bv. *viciae* strain WBAV_nodC_10, partial cds
CGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCATGTGCGATGTACCGTCGGTCTGCTATGCTTTCGCTGC
TCGATCAGTACGAGACGCAGCTTTATCGCGGCAAGCCGAGTGACTTCGGTGAAGATCGCCATTTGACGAT
TCTCATGCTGAGCGCCGGCTTTCGAACTGAGTATGTTCCGAGCGCCATCGCGGCGACAGTCGTTCCTGAC
ACAATGGGTGTTTATTTGCGTCAACAACTACGGTGGGCACGCAGCACCTTTCGGGATACTTTGCTTGCGC
TCCCCATACTGCCTGGCCTCGATCGGTATCTCACTCTGGACGTAATCGGGCAAAATGGCGGCCTTCTGCT
TCTTGCGCTGTCGGTATTGACGGGTATTGGCCAGTTTGCGCTGACCGCCACAGTACCATGGTGGACGATC
CTGGTGATCGGATCCATGACTCTTGTACGATGCAGCGTGGCTGCCTATCGCGCCCGCGAACTTAGGTTTC
TGGGTTTTGCTCTCCACACGCTCGTGAACATCTTTCTCTTAATTCCCTTGAAGGCCTATGCCCTTTGTAC
CTTGTC

>KP754737.1 *Rhizobium leguminosarum* bv. *viciae* strain WBAV_nodC_12, partial cds
CGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCATGTGCGATGTACCGTCGGTCTGCTATGCTTTCGCTGC
TCGATCAGTACGAGACGCAGCTTTATCGCGGCAAGCCGAGTGACTTCGGTGAAGATCGCCATTTGACGAT
TCTCATGCTGAGCGCAGGCTTTCGAACTGAGTATGTTCCGAGTGCCATAGCGGCGACAGTCGTTCCTGAC
ACAATGGGTGTTTATCTGCGTCAACAACTACGGTGGGCACGCAGCACTTTTCGGGATACATTGCTTGCGC
TCCCCATACTGCCTGGCCTCGATCGGTATCTCACTCTGGACGTAATCGGGCAAAATGGCGGCCTTCTGCT
TCTTGCGCTGTCGGTATTGACGGGTATTGGCCAGTTTGCGCTGACCGCCACAGTACCATGGTGGACGATC
CTGGTGATCGGATCCATGACTCTTGTACGATGCAGCGTGGCTGCCTATCGCGCCCGCGAACTTAGGTTTC
TGGGTTTTGCTCTCCACACGCTCGTGAACATCTTTCTCTTAATTCCCTTGAAGGCCTATGCCCTTTGTAC
CTTGTC

>EU697993.1 *Mesorhizobium* sp. CCNWNX0080 NodC (nodC) gene, partial cds
CAAAGCGCAGATCGCCGCGATACGCCGCTCATCTGGAGATTTGGTGCTCAACGTCGACTCGGACACGACA
CTCGCGTCCGACGTCATCACGAAGCTTGCACTGAAGATGCAGGATCCAGCCATCGGCGCTGCCATGGGCC
AGTTGACGGCAAGCAACCGGAGCGACACTTGGTTGACCCGGTTGATCGATATGGAGTACTGGCTCGCTTG
CAACGAGGAGCGTGCGGCACAGGCTCGCTTCGGTGCCGTTATGTGCTGTTGCGGCCCATGTGCTATGTAC
CGCCGGTCCTCGCTCGTTTCGCTGCTGGATCAGTACGAGACGCAGCTGTTTCGGGGGAAGCCAAGCGACT
TCGGTGAGGACCGCCATCTTACGATCCTCATGCTGAAAGCAGGCTTTCGAACCGAGTACGTTCCGGACGC
CATCGCGGCAACAGTCGTTCCGGACAGACTAGAGCC

>EU697986.1 *Phyllobacterium* sp. CCNWNX0083 NodC (nodC) gene, partial cds
GCAAAGCGCAGATCGCCGCGATACGCCGCTCATCTGGAGATTTGGTGCTCAACGTCGACTCGGACACGAC
AATCGCGTCCGACGTCATCACGAAGCTTGCACTGAAGATGCAGGATCCAGCCATCGGCGCTGCCATGGGC
CAGTTGACGGCAAGCAACCGGAGCGACACTTGGTTGACCCGGTTGATCGATATGGAGTACTGGCTCGCTT
GCAACGAGGAGCGTGCGGCACAGGCTCGCTTCGGTGCCGTTATGTGCTGTTGCGGCCCATGTGCTATGTA
CCGCCGGTCCTCGCTCGTTTCGCTGCTGGATCAGTACGAGACGCAGCTGTTTCGGGGGAAGCCAAGCGAC
TTCGGTGAGGACCGCCATCTTACGATCCTCATGCTAAAAGCAGGCTTTCGAACCGAGTACGTTCCGGACG
CCATCGCGGCAACAGTCGTTCCGGACAGACTAGAGCC

>EU697983.1 *Phyllobacterium* sp. CCNWNX0060 NodC (nodC) gene, partial cds
GCAGATCGCCGCGATACGCCGCTCATCTGGAGATTTGGTGCTCAACGTCGACTCGGACACGACAATCGCG
TCCGACGTCATCACGAAGCTTGCACTGAAGATGCAGGATCCAGCCATCGGCGCTGCCATGGGCCAGTTGA
CGGCAAGCAACCGGAGCGACACTTGGTTGACCCGGTTGATCGATATGGAGTACTGGCTCGCTTGCAACGA
GGAGCGTGCGGCACAGGCTCGCTTCGGTGCCGTTATGTGCTGTTGCGGCCCATGTGCTATGTACCGCCGG
TCCTCGCTCGTTTCGCTGCTGGATCAGTACGAGACGCAGCTGTTTCGGGGGAAGCCAAGCGACTTCGGTG
AGGACCGCCATCTTACGATCCTCATGCTAAAAGCAGGCTTTCGAACCGAGTACGTTCCGGACGCCATCGC
GGCAACAGTCGTTCCGGAC

>KP256196.1 *Rhizobium etli* bv. *trifolii* strain ECRI 31C NodC (nodC) gene, partial cds
GACGGTTCCAGAAATTGCGACGCCGTGGTCTCCGAGCAGATCGTCTATGCGGACGACTCGAGATTCGAAT
TCATCATGCTTCCCAGGAATGTCGGAAAGCGCAAAGCTCAAATCGCTGTCATTTCCCGATCGTCCGGGGA
CCTGATCTTGAACGTGGATTCATACATCACTCTCGCTTCCGACGTGGTTTCAAAACTCAGTCAGAAAATG
CGCGATCCGGCGGTCGGTGCGGTGATGGGGCAGCTTGTAGCAAGCAATCAGAGCGACTCGTGGCTGACCC
GATTGATCGACATGGAGTACTGGCTGGCCTGCAATGAGGAGCGCGCTGCGCAGGGCCGCTTCGGTGCCGT
GATGTGTTGCTGTGGACCCTGTGCTATGTACCGCCGGTCCGCCTTCGTCTTGCTTCTTGATCAATACGAG
ACGCAGCTTTATCGGGGAAAGCCGAGTGACTTTGGCGAGGACCGTCATCTGACCATCCTAATGCTAAGCG
CCGGCTTCCGTACCGAATATGTCCCAAGCGCCATCGCAGCAACCGTCGTTCCTGACGGTTTGGCTGCCTA
TCTGCGTCAGCAACTGCGTTGGGCACGGAGTACATTCCGAGATACCCTGCTTGGGCTTCACCTCCTCCGT
GGCATGAACTGGTATCTGACTTTGGACGTCATCGGGCAGAATGCCGGCCCTCTTCTGCTCGCATTATCCG
TACTGGCGGGTCTTGCACAGTTCGCCCTGACGGGTTCAGTGCCCTGGTGGTTGATAGGAGCGTTATTATC
ATTGACACTGTTACGATGCGGCGTCGATGTTTATCGTGCCAAGCAGCTTAGATTTCTTGG

>KP256193.1 *Rhizobium etli* bv. *trifolii* strain ECRI 21B NodC (nodC) gene, partial cds
AAACTCAGTCAGAAAAAGCGCGATCCTGCGGTCGGTGCGGTGATGGGGCAGCTTTTAGCAAGCAATTTGA
GCGAGTCGGGGTTGACCCGATTGATCGACATGGAGTTCTGGCTGGCCTGCAATGAGGAGCGCGCTGCGCA
GGGCCGCTTCGGTGCCGTGATGTGTTGCTGTGGACCCTGTGCTATGTACCGCCGGTCCGCCTTCGTCTTG
CTTTTTGATCAATACGTGATGCAGCTTTTTCGTGGAAAGTTGAGTTATTTTGGCGAGGACCGTCATCTGA
CCATCCTAATGCTAAGCGCCGGCTTCCGTACCGAATATGTTCCAAGCGCCATCGCAGCAACCGTCGTTCC
TGACGGTTTGGCTGCCTATCTGCGTCAGCAACTGCGTTGGGCACGGAGTACATTCCGAGATACCCTGCTT
GGGCTTCACCTCCTCCGTGGCATGAACTGGTATCTGACTTTGGACGTCATCGGGCAGAATGCCGGCCCTC
TTCTGCTCGCATTATCCGTACTGGCGTGTCTTGCACAGTTCGCCCCGACGGGTTCACTGCCCTGTTGGTT
GATAGGAACGATATTATCATTGACACTTTTATTATTCGG

>HQ231510.1 *Sinorhizobium fredii* strain CCBAU 23314 NodC (nodC) gene, partial cds
TGCTGCGGCCCGTGTGCCATGTACCGGCGGTCCGCACTCCTATTGCTGCTCGATAAATACGAGACGCAAC
TGTTTCGAGGCAGGCCAAGCGACTTCGGGGAAGACCGCCACCTCACAATCCTCATGCTGAATGCAGGCTT

TCGAACCGAGTACGTTCCGGACGCCATCGCGGCGACGGTCGTTCCAAACTCGATGGGGGCCTATCTGCGC
CAACAACTGCGCTGGGCACGCAGCACGTTTCGCGACACATTGCTCGCGCTCCGCCTACTGCCGGGCCTTG
ATCGCTATCTTACGCTGGACGTGATCGGACAGAATCTTGGTCCGCTGCTCCTAGCCCTCTCGGTCCTGAC
GGGGCTAGCACAGCTCGCTCTGACGGCCACAGTGCCTTGGTCGACGATCCTGATGATTGCATCTATGACA
ATGGTCCGCTGCGGCGTGGCGGCGTTTCGAGCGCGAGAGCTGCGATTCCTTGGGTTTTCGCTGCACACCC
TCCTCAACGTCGCTCTCCTGCTCCCCCTCAAAGCATATGCGTTGTGCACGTTGAGCAACAGCG

>HQ231505.1 *Sinorhizobium fredii* strain CCBAU 23319 NodC (nodC) gene, partial cds
TGCTGCGGCCCGTGTGCCATGTACCGGCGGTCCGCACTCCTATTGCTGCTCGATAAATACGAGACGCAAC
TGTTTCGAGGCAGGCCAAGCGACTTCGGGGAAGACCGCCACCTCACAATCCTCATGCTGAATGCAGGCTT
TCGAACCGAGTACGTTCCGGACGCCATCGCGGCGACGGTCGTTCCAAACTCGATGGGGGCCTATCTGCGC
CAACAACTGCGCTGGGCACGCAGCACGTTTCGCGACACATTGCTCGCGCTCCGCCTACTGCCGGGCCTTG
ATCGCTATCTTACGCTGGACGTGATCGGACAGAATCTTGGTCCGCTGCTCCTAGCCCTCTCGGTCCTGAC
GGGGCTAGCACAGCTCGCTCTGACGGCCACAGTGCCTTGGTCGACGATCCTGATGATTGCATCTATGACA
ATGGTCCGCTGCGGCGTGGCGGCGTTTCGAGCGCGAGAGCTGCGATTCCTTGGGTTTTCGCTGCACACCC
TCCTCAACGTCGCTCTCCTGCTCCCCCTCAAAGCATATGCGTTGTGCACGTTGAGCAACAGCG

>EU333392.1 *Bradyrhizobium japonicum* isolate RLA11 NodC (nodC) gene, partial cds
ATCTCGTGCTCAATGTCGATTCCGACACGGAAATCGCCCCCGACGTGGTCAGCAAGCTCGTATGCAAGAT
GCAGGATCCAGCCGTCGGCGCGGCCATGGGCCAATTGACGGCCAGCAACCGGAACGACAGCTGGCTGACC
CGCCTGATCGACATGGAGTACTGGTTGGCTTGCAACGAGGAACGCGCTGCGCAGACGCGCTTTGGTGCCG
TCATGTGCTGCTGTGGGCCATGTGCCATGTACCGCCGTTCCGCGCTCATGTTGCTGCTCGACCAATACGA
GACGCAGTTCTTCCGGGGAAAGCCGAGCGATTTCGGTGAGGACCGTCATCTGACCATCCTCATGCTCAAG
GCGGGGTTTCGCACCGAGTACGTTCCGGACGCCATCGCCGCAACGGTGGTCCCGGATAGGCTCTTGCCCT
ATCTGCGCCAACAGCTCCGCTGGGCGCGGAGCACCTACCGTGACACGCTGCTTGGTTTGCACCTGCTGCC
CGGCCTCGATCGGTACCTCACGCTCGATGTGCTCGGACAGAACCTCGGGCCGCTGCTGCTCGCCATCTCA
TCGATTGCCGCGCTCGCACAGCTCGCGCTCACGGACAGCGTGCCGTGGTGGACCGGTCTGACCATCGTCG
CGATGACCATGATTCGAT

>EU123541.1 *Sinorhizobium medicae* strain RPA18 NodC (nodC) gene, partial cds
ATTCGCGCGATCCGAGGTTCAGCTTCATTCTGCTCCCAGAGAACGTCGGAAAGCGGAAAGCGCAGATTGC
CGCGATAGGTCAATCCTCTGGGGATTTGGTGCTGAATGTCGACTCGGACAGCACGATCGCTTTCGATGTG
GTCTCCAAGCTTGCCTCGAAGATGGGAGATCCAGAGGTCGGTGCGGTTATGGGTCAACTCACGGCTAGCA
ATTCGGGTGACACTTGGCTGACGAAATTGATCGACATGGAGTATTGGCTTGCCTGCAACGAAGAACGCGC
GGCACAGGCTCGCTTCGGTGCTGTTATGTGTTGCTGCGGCCCTTGTGCTATGTACCGTCGGTCGGCGCTC
GCTTCGCTGCTTGACCAGTACGAAACGCAACTGTTTCGCGGTAAGCTAAGCGACTTCGGTGAGGACCGCC
ATCTGACGATCCTCATGTTGAAGGCAGGTTTTCGAACTGAGTATGTTCCAAACGCCATAGTGGCAACCGT
CGTCCCGGATACACTGAAACCGTATCTGCGCCAACAACTGCGTTGGGCACGCAGCACGTTCCGTGACACG
TTTCTAGTGCTCCCTCTGTTGCGCGGCCTCAACCCTTTTCTCACATTGGACGTGGTCGGGCAGAATATCG
GGCCACTGTTGCTCGCTCTGTCGGTCGTGACGGGACTTGCGCATTTCATAATGACCGCCACAGTGCCATG
GTGGACGATTTTGATTATTGCGTCCATGACCATTATACGCTGCAGCGTCGTAGCATTGCATGCTCGCCAA
CTTAGATTTCTTGGCTTCGTTCTGCACACACCCATCAACCTCTTTCTCTTACTTCCGTTGAAAGCTTATG
CGT

>EF694757.1 *Bradyrhizobium japonicum* strain BGA-1 NodC (nodC) gene, partial cds
CCCCGACGTGGTCAGCAAGCTCGTACGCAAGATGCAGGATCCAGCCGTCGGCGCGGCCATGGGTCAATTG

ACGGCCAGTAACCGGAACGACAGCTGGCTGACCCGATTGATCGACATGGAGTACCGGTTGGCTTGCAACG
AGGANCGCGCGGCTCAGACGCGCTTTGGTGCCGTCATGTGCTGCTGCGGGCCGTGTGCCATGTACCGCCG
TTCCGCGCTCATGTTGCTGCTCGACCAGTACGAGACGCAGTTCTTCCGGGGGAAGCCGAGCGATTTCGGT
GAGGACCGCCATCTGACCATCCTCATGCTCAAGGCAGGGTTTCGAACCGAGTACGTTCCGGACGCCATCG
CCGCCACCGTGGTCCCGGANAAGCTCTTGCCCTATCTGCGNCAACAGCTCCGCTGGGCGCGGAGCACCTA
CCGTGACACGCTGCTTGGTTTGCACCTGCTGCCCGGCCTCGATCGGTACCTCACGCTCGATGTGCTCGGA
CAGAACCTCGGGCCGCTGCTGCTCGCCATCTCATCGATTGCCGCGCTCGCACAGCTTGCGCTCACGGACA
GCGTGCCNTGGTGGACCGGTCTGACCATCGTCGCGATGACCATGATTCGATGC

>EF694752.1 *Bradyrhizobium canariense* strain MCLA12 NodC (nodC) gene, partial cds
ACTCCGACACGGAAATCGACAGCGACGTGGTCAGCAAGCTCGCGCGCAAGATGCAGGATCCAGCCGTCGG
CGCGGCCATGGGCCAATTGAAGGCCAGCAACCGAAACGACAGCTGGCTGACCCGATTGATCGACATGGAG
TACTGGTTGGCTTGCAACGAGGAGCGCGCGGCGCAGACGCGGTTTGGTGCCGTCATGTGMTGCTGTGGG
CGTGCGCCATGTACCGGCGATCCGCGCTGCTGTTGCTGCTCGAGCAATACGAGACGCAGTTCTTCCGGGG
AAAACCCAGCGACTTCGGTGAGGACCGTCATCTGACCATCCTCATGCTCAAGGCAGGGTTCCGAACGGAG
TACGTTCCGGACGCCATCGCCGCCACGGTGGTCCCGGACACGCTCTTGCCGTATCTGCGCCAACAGCTCC
GCTGGGCACGCAGCACCTACCGCGACACGCTGCTCGGTTTGCACCTGCTGCCCGGCCTCGACCGGTACCT
CACGCTGGACGTGGTTGGACAGAACCTCGGCCCGCTGCTGCTCGCCATCTCATCGANTGCCGCGCTCGCT
CAGCTCGCGCTCACGGGCTCCGTGCCCTGGTGGACCGGTCTGACCATCGTCGCGATGACCA

>DQ413013.1 *Rhizobium tropici* isolate RP261 NodC (nodC) gene, partial cds
GATCCAAGATTCACATTCATTCTGCTCAGGGAGAATGTTGGAAAGCGCAAGGCGCAGATAGCAGCGATAC
GCCGTTCCTCTGGGGATTTGGTCCTCAACGTCGACTCCGACACCATTCTCGCCCCAGACGTCGTCGTGAA
ACTTGCACTCAAGATGCAAGATCCGGCAATTGGCGCGGCTATGGGCCAGCTTGCCGCCAGCAACCGCCAC
GAGACCTGGTTGACCCGTTTGATCGACATGGAATACTGGCTGGCCTGTAATGAGGAGCGCGCGGCACAGG
CTCGTTTCGGTGCTGTCATGTGCTGCTGTGGCCCATGCGCCATGTACCGCCGGACTGCGCTTACCATGCT
ACTCGACCAGTACGAAACGCAAATGTTCCGGGGAAAGCGAAGCGATTTCGGCGAAGATCGTCATCTTACG
ATTCTCATGTTGAAGGCCGGATTTCGAACCGAATATGTTCCAACCGCCATCGCGGCAACTGTTGTTCCAA
ACAAGCTGCGCCCTTATCTGCGTCAACAACTTCGCTGGGCACGCAGCACTTTCCGTGACACGTTGCTTGC
AATGAACCTTCTGCCTGGTCTTGACCGTTTTCTCACGTTGGACGTCATTGGTCAGAACCTCGGACCGCTC
TTGCTCGCCCTATCGGTGCTAACCGGACTAGCGCAGTTCGCGCTCACCGGGACCGTGCCCTGGTGGACAT
GCCTGATGATCGCATCGATGACCATGATCCGCTGCAGCGTCGCAGCGGTTCGTGCTCGCCAATTTCGATT
TATCGGCTTCTCCCTGCATACCTTCATCAAC

>DQ413011.1 *Rhizobium etli* bv. *phaseoli* isolate RP338 NodC (nodC) gene, partial cds
GACCCCAGATTCAATATTCTCCTGCTTCCCCAGAATGTCGGTAAACGGAAGGCACAGATCGCCGCGATAC
GCCGCTCTGCTGGAGATATGGTGTTAAACGTCGACTCCGACACGATCCTCGCATCTGACGTCATCAGGAA
GCTCGTGCCTAAAATGCAAGATCCGGCTGTCGGCGCGGCCATGGGACAGTTGACGGCCCGCAACCGAAAC
GATAGTTGGCTGACCCGTTTGATCGATATGGAGTACTGGCTGGCTTGCAACGAGGAGCGTGCGGCACAAG
CTCGCTTCGGAGCCGTTATGTGCTGCTGCGGTCCATGTGCTATCTACCGTCGCTCTGCGCTCGCTTCGCT
GCTTGACCAGTACGAATCACAGTATTTTCGGGGAAAGCCAAGCGATTTCGGTGAGGATCGGCATCTCACC
ATTCTCATGCTGAAGGCAGGCTTTCGAACGGAGTACGTGCCGAGCGCCATCGCAGCGACAGTCGTTCCGA
ACAAGCTAGGACCGTATCTGCGCCAACAACTACGCTGGGCGCGGAGCACGTTCCGGGACACGTTGCTTGG
GCTGCGCCTGCTGCCCAACCTCAATCGCTTCCTTACGCTCGACGTTGTCGGACAGAACCTCGGACCGCTG
CTTCTGGCACTATCAGTGCTGACGGGGCTCGCACAGCTTGCATTGACGGGCAACGTGCCTTGGTTGCCAG
CCCTGATGATTGTGGCCATGACGATGATCCACTGCAGCGTTGTCGCGCTTCGGGCCCGCCAACTACGGTT

CCTCGGGTTCTCTCTGCACGCATTTATCAGT

>DQ376607.1 *Bradyrhizobium elkanii* isolate 25.16 NodC (nodC) gene, partial cds
GTCGATTGCCGGTCAAGACTACGCCGGAAAACTGAGAGTCATTGTGGTCGACGATGGATCCGCAAATCGC
GATCTGTTGGGACCCGTACACAAAATCTATGCGAGCGATCCGAGGTTTCGCATCATCTTGATGGCCAAGA
ACGTCGGAAAGCGCAAGGCGCAGATTGCTGCGATACGCAGCTCTTCCGGCGATCTGGTCCTGAACGTCGA
CTCGGATACGATCCTTGCCGTCGACGTGGTCACGAAGCTTGTATCGAAGATGCAGGACCCCGACGTCGGC
GCGGCCATGGGGCAGTTGGTAGCGAGCAATCGCAACCAGACCTGGC

>EF201794.1 *Sinorhizobium medicae* strain RPA11 NodC (nodC) gene, partial cds
GATCCGAGGTTCAGCTTCATTCTGCTCCCAGAGAACGTCGGAAAGCGGAAAGCGCAGATTGCCGCGATAG
GTCAATCCTCTGGGGATTTGGTGCTGAATGTCGACTCGGACAGCACGATCGCTTTCGATGTGGTCTCCAA
GCTTGCCTCGAAGATGGGAGATCCAGAGGTCGGTGCGGTTATGGGTCAACTCACGGCTAGCAATTCGGGT
GACACTTGGCTGACGAAATTGATCGACATGGAGTATTGGCTTGCCTGCAACGAAGAACGCGCGGCACAGG
CTCGCTTCGGTGCTGTTATGTGTTGCTGCGGCCCTTGTGCTATGTACCGTCGGTCGGCGCTCGCTTCGCT
GCTTGACCAGTACGAAACGCAACTGTTTCGCGGTAAGCTAAGCGACTTCGGTGAGGACCGCCATCTGACG
ATCCTCATGTTGAAGGCAGGTTTTCGAACTGAGTATGTTCCAAACGCCATAGTGGCAACCGTCGTCCCGG
ATACACTGAAACCGTATCTGCGCCAACAACTGCGTTGGGCACGCAGCACGTTCCGTGACACGTTTCTAGT
GCTCCCTCTGTTGCGCGGCCTCAACCCTTTTCTCACATTGGACGTGGTCGGGCAGAATATCGGGCCACTG
TTGCTCGCTCTGTCGGTCGTGACGGGACTTGCGCATTTCATAATGACCGCCACAGTGCCATGGTGGACGA
TTTTGATTATTGCGTCCATGACCATTATACGCTGCAGCGTCGTAGCATTGCATGCTCGCCAACTTAGATT
TCTTGGCTTCGTTCTGCACACACCCATCAACCTCTTTCTCTTAC

>KU664565.1 *Rhizobium bangladeshense* strain 1017 NodC (nodC) gene, partial cds
CAGCTTGCCTATGCGGGCGACGCGAGATTCGAATTCATCATGCTTCCCAGGAATGTCGGAAAGCGCAAAG
CGCAAATCGCGGCCATTTCCCGATCATCCGGGGACCTGATCCTGAACGTCGATTCAGACACCACTCTCGC
CAGCGATGTTGTTTCCAAACTCAGTCAGAAAATGCGCGATCCGACGGTCGGTGCGGTGATGGGGCAGCTT
GTAGCAAGCAATCAGAGCGACTCGTGGCTGACCCGATTGATCGACATGGAGTACTGGCTGGCCTGCAATG
AGGAGCGCGCCGCGCAGGGCCGCTTCGGTGCCGTGATGTGTTGCTGTGGACCTTGTGCCATGTACCGCCG
GTCCGCCTTCGTCTTGCTTCTTGATCAATACGAGACGCAGCTTTATCGGGGAAAGCCGAGTGACTTTGGC
GAGGACCGTCATCTGACCATCCTAATGCTAAGCGCCGGCTTCCGTACCGAATATGTCCCAAGCGCCATCG
CCGCAACCGTCGTTCCTGACGGTTTGGCTGCCTATCTGCGCCAGCAACTGCGTTGGGCACGGAGTACATT
CCGAGATACCCTGCTTGGGCTTCACCTCCTCCGTGGCATGAACTGGTATTTGACTTTGGACGTCGTCGGG
CAGAATGCCGGCCCTCTTCTGCTCGCATTATCCGTACTGGCGGGTCTTGCACAGTTCGTCCTGACGGGTT
CAGTGCCCTGGTGGACGATAGGAACGATTGGATTGTTGACACTGATACGATGCGGCGTAGCTGCTTATCG
TGCCAAGCAGCTTAGATTTCTTGGCTTCTCGCTGCACACGCTCGTGAACGTCTTTCTATTACTTCCGGTG
AAGGCT

>KT718954.1 *Bradyrhizobium* sp. 982N22 NodC (nodC) gene, partial cds
GCGGCCCATGCGCGATGTATCGTCGCTCTGCGCTCGCCTTGCTGCTTGATCAATACGAAGCGCAGTTCTT
TCGTGGAAAGCCGAGCGATTTCGGCGAGGATCGCCATCTAACGATCCTCATGCTCAAGGCGGGGTTTCGA
ACCGAATATGTTCCGGACGCGATCGCAGCGACTGTCGTCCCGGACAGCCTTGGGCCATATCTGCGTCAGC
AACTCCGCTGGGCGCGCAGTACCTTTCGGGACACGTTTCTTGCACTGCGCCTGCTGCCGGAGCTCGATGG
CTATCTGACGCTGGACGTTGTCGGGCAGAATCTCGGCCCGTTGCTTCTCGCCCTTTCGTCGCTGGCTGCG
CTCGCACAGCTCGTAGTCGGCGGTTCGGTACCCTGGTGGACGGGACTGACGATCGCAGCAATGACGATGA
TCCGGTGCAGTGTGGCAGCCCTTCGTGCTCGCGAGTGGCGGTTTCTCGGCTTCTCGCTCCACACACCGAT

CAATATCTTTCTCTTACTACCTCTGAAAGCCTATGCGCTTTGTACATTGAGCAA

>KT718888.1 *Burkholderia* sp. 976N8 NodC (nodC) gene, partial cds
GTGGTCCGTGTGCCATGTACCGCCGGTCCGCCCTCCTTTTGCTGCTCGATCAGTACGAGACGCAACTTTT
TCGCGGTAAACCGAGCGACTTCGGCGAGGACCGCCATCTCACGATTCTCATGCTGACGGCAGGTTATCGG
ACCGAGTACGTTCCGGATGCCATTGCAGCGACAGTCGTTCCGGACAGACTGGGGGCGTATCTGCGCCAAC
AACTGCGCTGGGCGCGCAGTACTTATCGGGACACGTTGCTTTCGCTGCGCCTGCTGCCCTGTCTAGATCG
CTACCTTACGCTGGACGTGATCGGGCTGAATCTGGGCTCGCTATTTCTCGCCTTATCGCTTCTAGCGGCG
CTCGCGCAGCTCGCATTAACAGCGACAGTGCCGTGGTGGACTGCCCTGATAATCGCATCATGGACCATGA
TCCGCTGTAGCGTGGCATCGGTTCGCGCTCGCCAGATTCGATTTCTTGGCTTTTTTCTTCACACACCCAT
CAACCTCTTTCTCTTACTTCCCCTGAAAGCCTATGCGCTGTGTACATTGAGCAACAGC

>KT718883.1 *Burkholderia* sp. 976N3 NodC (nodC) gene, partial cds
GCGGTCCATGTGCCATGTACCGCCGGTCCGCGCTCCTTTTGCTGCTAGATCAGTACGAGACGCAATTTTT
TCGCGGGAAGCCGAGCGACTTCGGCGAGGACCGTCATCTGACGATTCTCATGCTGACGGCAGGTTATCGG
ACCGTGTACGTTCACGACGCCATTGCGGCGACAGTCGTTCCAGACAGACTGGGGGCGTATCTGCGCCAAC
AACTGCGCTGGGCGCGCAGTACGTATCGGGACACGTTGCTTTCGCTGCGCCTGCTGCCCCGTCTCGATCG
TTACCTTACGCTGGACGTGATCGGACACAATCTGGGCTCGCTATTTCTCGGCTTATCGCTACTAGCGGGT
CTCGCGCAGCTCGCATTGACAGCGACAGTGCCGTGGTGGACTGCCCTGATAATCGCATCATCGACCATGA
TCCGCTGTAGCGTGGCATCGGTTCGCGCTCGCCAGGTTCGATTTCTTGGCTTTTCTCTTCACACACCCAT
TAACCTCTTTCTCTTGCTCCCCCTGAAAGTCTATGCGCTGTGTACATTGAGCAACAGC

>KT718862.1 *Burkholderia* sp. 975N5 NodC (nodC) gene, partial cds
GTGGTCCGTGTGCCATGTACCGCCGGTCCGCGCTCCTGTTGCTGCTCGATCAGTACGAGACGCAACTTTT
TCGCGGGAAACCGAGCGACTTCGGCGAGGACCGCCATCTCACGATTCTCATGCTGACGGCAGGTTATCGG
ACCGAGTACGTTCCGGATGCCATTGCGGCGACAGTCGTTCCAGACAGACTGGGGGCGTATCTGCGCCAAC
AACTGCGCTGGGCGCGCAGTACTTATCGGGACACGTTGCTTTCGCTACGCCTGCTGCCCTGTCTCGATCG
CTACCTTACGCTGGACGTGATCGGACTGAATCTGGGCTCGCTATTTCTCGCCTTATCGCTTCTAGCGGGG
CTCGCGCAGCTCGCATTAACAGCGACAGTGCCGTGGTGGACTGCCCTGATAATCGCATCATGGACCATGA
TCCGCTGTAGCGTGGCATCGGTTCGCGCTCGCCAGATTCGATTTCTTGGCTTTTCTCTTCACACACCCAT
CAACCTCTTTCTCTTGCTCCCCCTGAAAGCCTATGCGCTGTGTACATTGAGCAACAGC

>KP308194.1 *Rhizobium giardinii* strain S3 NodC (nodC) gene, partial cds
GTGCAGGCCGTTTATTCTCTGCCGACCGATGTTTCATTGGCGTCCCAGAGCTTGGCCGGCTTTGAGGAGC
TGCCCAGCGTAGATGTCATCGTGCCAAGCTTCAACGAGGATCCCCGCACGCTTTCGGAGTGCCTGGCTTC
TATTGCGGGTCAGGAATACGGGGGAAGGCTGCAGGTTTACCTAGTTGATGACGGTTCCGAAAATCGCGAG
GCTTTGCGACCTGTGCACGAGGCCTTCGCACGAGACCCCAGATTCAATATTCTCCTGCTTCCCCAGAATG
TTGGTAAACGGAAGGCACAGATCGCCGCGATACGCCGCTCTGCTGGAGATATGGTGTTAAACGTCGACTC
CGACACGATCCTCGCATCTGACGTCATCAGGAAGCTCGTGCCTAAAATGCAAGATCCGGCTGTCGGCGCG
GCCATGGGACAGTTGACGGCCCGCAACCGAAACGATAGTTGGCTGACCCGTTTGATCGATATGGAGTACT

>KP308192.1 *Rhizobium leguminosarum* bv. *phaseoli* strain HM1 NodC (nodC) gene, partial cds
GCGGTTGTGGCTCAGCGCGCTGTCTATGCAGACGATGAGAGATTCAACTTCACAATTCTCCCTAAAAATG
TTGGAAAGCGCAAAGCGCAAATCGCCGCTATAACCCAGTCCTCTGGGGATCTCATTTTGAATGTGGACTC
AGACACCACGATCGCCCCCGACGTCGTATCTAAGCTTGCCCACAAAATGCGGGATCCAGCAGTCGGTGCG
GCGATGGGCCAAATGAAAGCCAGCAACCAGGCGGACACCTGGCTAACTCGCTTGATTGACATGGAGTACT

GGCTTGCCTGCAACGAGGAGCGCGCGGCACAAGCTCGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCATG
TGCGATGTACCGTCGGTCTGCTATGCTTTCGCTGCTCGATCAGTACGAGACGCAGCTTTATCGCGGCAAG
CCGAGTGACTTCGGCGAAGATCGCCATTTGACGATTCTCATGCTGAGCGCAGGCTTTCGAACTGAGTATG

>KC768861.1 *Microvirga vignae* strain BR3299 NodC (nodC) gene, partial cds
GGCTGGCTTGCAACGAGGAGCGCGCGGCACAGGCTCGCTTCGGTGCCGTTATGTGCTGCTGCGGTCCATG
TGCTATGTACCGTCGGGCCGCGCTTCTTTCGCTGCTAGACGAGTACGAGACGCAGTTGTTTCGTGGGAAG
CCAAGCGACTTCGGCGAGGATCGCCATCTCACGATCCTCATGTTGAAAGCAGGGTTTCGAACGGAGTACG
TTTCCGACGCAATTGCAGCAACAGTTGTTCCGGATAAGGTAGGGCCATATTTGCGCCAACAACTCCGCTG
GGCACGGAGCACCTTTCGCGACACGCTGCTTGCGCTGCGCCTGCTGCCAGCTCTTAATGGGTATCTCACA
CTAGATGTAATTGGGCAGAACCTCGGCCCGCTACTTCTCGCCCTGTCATCGCTAACGGCGCTCGCGCAGC
TCGCGCTCACCGCCACAGTTCCGTGGTGGACAGCATTGGTGATCGCGTTGATGACTTTCGTTCGCTGCAG
CGTGGCGGCAATTCGTGCTCGCCAAATTCGATTTCTCGGCTTTGCCCTGCACAC

>FJ596028.1 *Rhizobium leguminosarum* strain PEVF05 NodC (nodC) gene, partial cds
ATGCAGACGATGAGAGATTCAAATTCACAATTCTCCCTAAAAATGTTGGAAAGCGCAAAGCGCAAATCGC
CGCTATAACCCAGTCCTCTGGGGACCTCATCTTGAATGTGGACTCAGACACCACGATCGCCCCCGACGTC
GTCTCGAAGCTTGCCCATAAAATGCGCGATCCAGCAGTCGGTGCGGCGATGGGCCAAATGAAAGCCAGTA
ACCAGGCGGACACCTGGCTAACTCGCTTGATTGACATGGAGTACTGGCTTGCCTGCAATGAGGAGCGCGC
GGCGCAAGCTCGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCATGTGCGATGTACCGTCGGTCTGCTATG
CTTTCGCTGCTCGATCAGTACGAGACGCAGCTTTATCGCGGCAAGCCGAGTGACTTCGGTGAAGATCGCC
ATTTGACGATTCTCATGCTGAGCGCAGGCTTTCGAACTGAGTATGTTCCGAGCGCCATCGCGGCGACAGT
CGTTCCAGACACAATGGGAGTTTATCTGCGTCAACAACTACGGTGGGCACGCAGCACTTTTCGGGATACA
TTGCTTGTGCTTCCCGTACTACCTGGTCTCGATCGGTATCTCACGCTGGACGCAATCGGGCAAAATGTCG
GCCTTCTACTTCTTGCGCTGTCGGTATTGACAGGAATTGGTCAGTTTGCGCTGACCGCCACAGTACCCTG
GTGGACGATCCTGGTCATTGGATCCATGACTCTTGTACGGTGTAGCGTGGCTGCCTATCGCGCCCGCGAA
CTTAGGTTTTTGGGTTTTGCTCTCCACACGCTCCTGAACATCTTTCTCTTAA

>FJ596025.1 *Rhizobium leguminosarum* strain PEVF01 NodC (nodC) gene, partial cds
ATGCAGACGATGAGAGATTCAACTTCACAATTCTCCCTAAAAATGTTGGAAAGCGCAAAGCGCAAATCGC
CGCTATAACCCAGTCCTCTGGGGATCTCATTTTGAATGTGGACTCAGACACCACGATCGCCCCCGACGTC
GTATCTAAGCTTGCCCACAAAATGCGGGATCCAGCAGTCGGTGCGGCGATGGGCCAAATGAAAGCCAGCA
ACCAGGCGGACACCTGGCTAACTCGCTTGATTGACATGGAGTACTGGCTTGCCTGCAACGAGGAGCGCGC
GGCACAAGCTCGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCATGTGCGATGTACCGTCGGTCTGCTATG
CTTTCGCTGCTCGATCAGTACGAGACGCAGCTTTATCGCGGCAAGCCGAGTGACTTCGGCGAAGATCGCC
ATTTGACGATTCTCATGCTGAGCGCAGGCTTTCGAACTGAGTATGTTCCGAGTGCCATCGCGGCGACAGT
CGTTCCAGACACGATGGGTGTTTATCTACGCCAACAACTACGCTGGGCACGCAGCACCTTTCGGGATACT
TTGCTTGCGCTCCCCGTACTGCCTGGTCTCGATCGATATCTCACGCTGGACGCAATCGGGCAAAATGTCG
GCCTGCTACTTCTTGCGCTGTCGGTATTGACAGGAATTGGCCAGTTTGCGCTGACCGCCACACTACCCTG
GTGGACGATCCTGGTCATTGGATCCATGACTCTTGTCCGCTGTAGCGTGGCTGCCTATCGCGCCCGCGAA
CTGAGGTTTCTGGGTTTTGCTCTCCACACGCTTGTGAACATCTTTCTCTTAGTTCCCTTGAAGGCCTATG
CCCTTTGACC

>FJ596019.1 *Rhizobium leguminosarum* strain RVS11 NodC (nodC) gene, partial cds
TATGCAGACGATGAGAGATTCAAGTTCACAATTCTCCCTAAAAATGTGGGAAAGCGCAAAGCGCAAATCG
CCGCTATAACCCAGTCCTCTGGGGACCTCATCTTGAATGTGGACTCAGACACCACGATCGCCCCCGACGT

CGTCTCGAAGCTTGCCCATAAAATGCGCGATCCAGCAGTCGGTGCGGCGATGGGCCAAATGAAAGCCAGC
AACCAGGCGGACACCTGGCTAACTCGCTTGATTGACATGGAGTACTGGCTTGCCTGCAATGAGGAGCGCG
CGGCGCAAGCTCGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCATGTGCGATGTACCGTCGGTCTGCTAT
GCTTTCGCTGCTTGATCAGTACGAGACGCAGCTTTATCGCGGCAAGCCGAGTGACTTCGGCGAAGATCGC
CATTTGACGATTCTCATGCTGAGCGCAGGCTTTCGAACTGAGTATGTTCCGAGCGCCATCGCGGCGACAG
TCGTTCCAGACACAATGGGAGTTTATCTGCGTCAACAACTACGGTGGGCACGCAGCACTTTTCGGGATAC
TTTGCTTGCGCTTCCCGTACTACCAGGTCTCGATCGATATCTCACGCTGGACGCAATCGGGCAAAATGTC
GGCCTTTTACTTCTTGCGCTGTCGGTATTGACAGGAATTGGTCAGTTTGCGCTGACCGCCACAGTACCCT
GGTGGACGATCCTGGTCATTGGATCCATGACCCTTGTACGGTGTAGCGTGGCTGCCTATCGTGCCCGCGA
ACTTAGGTTTCTGGGTTTTGCTCTCCACACGCTCCTGAACATCTTTCTCTT

>KR154722.1 *Burkholderia* sp. BL27 I3R6 NodC (nodC) gene, partial cds
TCGCTTCGGTGCTGTTATGTGTTGCTGCGGTCCGTGTACCATATACCGACGGTCCGCGCTTCTTTTACTG
TTAGACCGGTACGAGACACAGCTGTTTCGGGGGAAACCGAGCGACTTCGGCGAGGACCGCCATCTCACGA
TCCTCATGCTGACGGCAGGGTTGCGGACCGTGTACGTCCCGGACGCCATTGCGGCGACAGTCGTTCCGGA
CAGGCTGGGGGCGTATCTGCGCCAACAGCTGCGCTGGGCGCGCAGTACTTATCGGGACACGTTGCTTGCA
CTGCGCCTGCTGCCTGACCTCAATCGCTACCTTACGCTGGACGTGATCGGACTGAATCTGGGCCCGTTAT
TGCTTGCACTGTCGTTTCTAACCGGACTCTTTCAGCTCGCGCTGACAGCCACGGCACCCTGGCAGGCAGG
CCTGATAATCGCATCATTGACCATGATCCGCTGTAGCGTGGCATCGGTTCGCGCTCGCCAGGTCCGATTT
CTTGGATTTTCCCTACACGCGCTAATCACTCTTTTCCTCTTGCTCCCCGTGAAGGCCTATGCCTTATGTA
CGTTGAGCAACAGC

>KR154717.1 *Burkholderia* sp. BL23 I1R1 NodC (nodC) gene, partial cds
TCGCTTCGGCGCTGTTATGTGTTGCTGTGGTCCGTGTGCCATGTACCGCCGGTCCGCGCTCCTTCTGCTG
CTAGATCAGTACGAGACGCAACTTTTTTGCGGGAAACCGAGCGACTTCGGCGAGGACCGCCATCTCACGA
TTCTCATGCTGACGGCAGGTTATCGGACCGAGTACGTTCCGGACAGTCTGGGGGCGTATCTGCGCCAACA
ACTGCGCTGGGCGCGCAGTACGTATCGGGACACGTTGCTTTCGCTGCGCCTGCTGCCCTGTCTCGATCGC
TACCTTACGCTGGAGGTGATCGGACTGAATCTGGGCTCGCTATTTCTCGCCTTATCGCTTCTAGCGGGGC
TCGCGCAGCTTGCATTGACAGCGACAGTGCCGTGGTGGACTGCCCTGATAATCGCATCATGGACCATGAT
CCGCTGTAGCGTGGCATCGGTTCGCGCTCGCCAGATTCGATTTCTTGGCTTTTCTCTTCACACACCCATT
AACCTCTTTCTCTTGCT

>KX943561.1 *Mesorhizobium* sp. CSLC37N N-acetylglucosaminyltransferase NodC (nodC) gene, partial cds
GTCGGAAAGCGCAAGGCGCAGATTGCCGCCATTCGCCGCTCGTGCGGAGATTTGGTGCTGAATGTAGATT
CGGATACGATACTCGCCCCGGACGTCATCACCAGGCTTGCGCTAAAGATGCAAGATCAAGCGGTCGGCGC
CGCCATGGGCCAGTTGGCGGCTAGCAATCGCAGTGAAACTTGGCTGACGCGGTTGATCGACATGGAATAC
TGGCTCGCCTGCAACGAAGAGCGGGCAGCGCAGGCTCGGTTCGGTGCCGTCATGTGTTGCTGCGGTCCGT
GTGCCATGTACCGGCGGTCTGCGCTTGTTTCGCTGCTGGATCAGTACGAGACGCAGCGCTTTCGAGGGAA
GCCGAGCGACTTCGGCGAGGACCGCCACCTTACGATCCTGATGCTGAAAGCAGGCTTTCGAACAGAGTAT
GTCCCGGAGGCCGTCGCGGCAACAGTCGTTCCCAACAGCATGGGTCCCTATCTGCGCCAACAGCTTCGCT
GGGCCCGGAGCACTTTCCGTGACACGTTGCTGGCGTTCCAACTGCTGCGCGGCCTTAATATTTATCTCAC
ATTGGACGTGATTGGCCAGAATATTGGCCCATTATTGCTCTCTTTGTCGATTCTGGCAGGGCTCGCGCAA
TTCGTAACGACAGGTACTGTGCCTTGGACGGCATGCCTGATGATTGCAGCCATGACTATAGTTCGCTGCA
GCGTGGCAGCGTTTCGTGCGCGCCAACTTCGATTTCTCGGATTCTCGCTCCACACACTCATCAACATCTT
TCTCTTGCTCCCGTTGAAAGCATAC

>EF549513.1 *Rhizobium* sp. CCBAU 83304 N-acetylglucosaminyltransferase (nodC) gene, partial cds

GCGTCTGAACAAACACGCTTCGGTGCCGTGATGTGTTGCTGTGGACCCTGTGCCATGTACCGCCGGTCCG
CCTTCGTCTTGCTTCTTGATCAATACGAGACGCAGCTTTATCGGGGAAAGCCGAGTGACTTTGGCGAGGA
CCGTCATCTGACCATTCTAATGCTAAGCGCCGGCTTCCGTACCGAATATGTCCCAAGCGCCATCGCCGCA
ACCGTCGTTCCTGACGGTTTGGCTGCCTATCTGCGCCAGCAACTGCGTTGGGCACGGAGTACATTCCGAG
ATACCCTGCTCGGACTTCACCTCCTCCGTGGCATGAACTGGTATCTGACTTTGGACGTCGTCGGGCAGAA
TGCCGGCCCTCTTCTGCTCGCTTTGTCCGTACTGGCGGGTCTTGCACAGTTTGCCTTGGCGGGTTCAGTG
CCCTGGTGGACGATAGGAACGATTGGATCATTGACACTGATACGATGCGGCGTAGCTGCTTATCGTGCCA
GGCAGCTTAGATTTGTTGGCTTCTCGCTGCACACGCTTGTGAACGTCTTTCTATTACTTCCTGTGAAAGC
TTATGCGCTCTGTACGTTGTCCAACAGGGATTGGTTTGGTCGGGGAC

>HQ588006.1 *Bradyrhizobium* sp. HM125 N-acetylglucosaminyltransferase (nodC) gene, partial cds

TGTTGGCGAACAATGTGGGAAAGCGCAAGGCGCAGATCGCAGCGATACGCAGCTCATCCGGTGATCTGGT
TCTCAACGTCGATTCCGATACGATACTTGCTGCCGACGTCGTCACGAAGCTTGTATTGAAGATGCATGAC
CCGGGAATCGGTGCGGCCATGGGTCAGCTGATCGCGAGCAATCGCAACCAGACCTGGCTGACCAGGCTGA
TCGACATGGAATATTGGCTCGCGTGCAACGAAGAGCGCGCGGCACAGGCGCGCTTCGGTGCCGTCATGTG
TTGCTGCGGCCCATGTGCCATGTATCGGCGTTCCGCGCTCGCCTTGCTTCTTGATCAATATGAAGCCCAA
TTCTTTCGTGGGAAGCCGAGCGATTTCGGCGAGGACCGCCACCTAACGATACTCATGCTCAAGGCGGGGT
TTCGAACCGAATACGTTCCGGACGCCATAGCAGCCACAGTCGTCCCGCACAGTCTTCGGCCATATCTACG
ACAGCAACTCCGCTGGGCGCGAAGTACCTTTCGAGATACGTTTCTTGCTTGGCGCCTGCTGCCAGAGCTC
GATGGTTATTTGACGCTAGACGTTATCGGGCAAAATCTCGGCCCATTGCTCCTCGCCATTTCATCACTTG
CTGCGCTCGCACAGCTCCTGATCGATGGCTCTATACCCTGGTGGACGGGATTGACGATTGCTGCAATGAC
TACGGTCCGGTGCTGTGTGGCAGCGCTT

>KC848542.1 *Sinorhizobium meliloti* strain STM3226 N-acetylglucosaminyltransferase (nodC) gene, partial cds

GGAACCGCGAGGCAATTGTGCGTGTACACGATTTCTATTCGCGCGATCCGAGGTTCAGCTTCATTCTGCT
CCCAGAGAACGTCGGAAAGCGGAAGGCGCAAATTGCCGCGATAGGTCAATCCTCTGGGGATTTGGTGCTG
AATGTCGACTCAGACAGCACGATTGCTTGCGATGTCGTCTCGAAGCTTGCCTCGAAGATGCGAAATCCAG
AGGTCGGTGCGGTCATGGGTCAACTCACGGCTAGCAATCGGAGTGACACCTGGCTGACGAAATTGATCGA
CATGGAGTATTGGCTTGCCTGCAACGAAGAACGCGCGGCACAGGCTCGCTTCGGTGCGGTTATGTGTTGC
TGCGGCCCATGTGCTATGTACCGCCGGTCGGCGCTCGTTTCGCTGCTGGACCAGTACGAAACGCAACTGT
TTCGCGGTAAACCAAGCAACTTCGGTGAGGATCGCCATCTGACGATCCTCATGTTGAAGGCAGGCTTTAG
AACTGAGTATGTTCCAGACGCCGTAGTGGCGACCGTCGTTCCAGATAGGCTGAAACCCTATCTGTGCCAA
CAACTGCGTTGGGCACGCAGCACGTTCCGCGACACGTTTCTAGCGCTCCCTCTGTTGCGAGGCCTCAACC
GTTTTATCACATTGGACGTGGTCGGGCAGAATGTGGGGCCGCTGTTGCTCGCCTTATCGGTAGTGACGGG
ACTTGCGCACTTCATCATGTCCGCCACAGTGCCGTGGTGGACAATTTTGATGATTGTATCCATGACCATG
ATACGCTGCAGCGTGGTAGCATTGCATTCTCGCCAACTTAGATTTCTTGGCTTCGTTCTGCACACACCCA
TCAACCTCTT

>KY056644.1 *Rhizobium leguminosarum* bv. *viciae* strain Mi2 N-acetylglucosaminyl transferase (nodC) gene, partial cds

CGCTGCCTATGCAGACGATGAAAGATTCAGCTTCATAATTCTCGCTAAAAATGTTGGAAAGCGCAAAGCG
CAAATCGCCGCTATCACCCAGTCCTCTGGGGATCTCATCTTGAACGTGGACTCAGACACCACGCTCGCCC

CCGACGTCGTCTCCAAGCTCGCCCACAAAATGGGCGATCCAGCGGTCGGTGCGGCGATGGGCCAATTGAA
AGCCAGTAACCAGGGGGACACTTGGCTGACTCGCTTGATTGACATGGAGTACTGGCTTGCCTGCAACGAG
GAGCGCGCGGCGCAAGCTCGCTTCGGTGCAGTTATGTGCTGCTGCGGCCCATGCGCGATGTACCGGCGTT
CTGCTATGCTCTCGCTGCTGGATCAGTACGAGACGCAACTTTATCGCGGCAAGCCGAGTGACTTCGGTGA
AGATCGCCATTTAACGATTCTCATGCTGAGCGCAGGTTTTCGAACTGAGTATGTTCCGAGCGCCATCGCG
GCGACAGTTGTTCCTGACACAATAGGTGTTTATCTGCGTCAACAACTACGGTGGGCACGCAGCACTTTTC
GGGATACTTTGCTTGCGTTCCCCGTACTGCCTGGCCTTGATCGATATCTCACGCTGGACGTAATCGGGCA
AAATGGCGGCCCTCTACTTCTTGCGCTGTCGGTGTTGACGGGTATTGGCCAGTTTGCGTTGACCGCCACA
GTACCATGGTGGACGATCATGGTCATCGTCTCCATGACTCTTGTGCGATGCAGTGTGGCTGCCTATCGCG
CCCGCGAGCTTAGGTTTCTGGGTTTTGCTCTCCACACGCTCGTGAACATCTA

>KU939441.1 *Rhizobium leguminosarum* bv. *viciae* strain LMR416 N-acetylglucosaminyltransferase (nodC) gene, partial cds
GTTGGAAAGCGCAAAGCGCAAATCGCCGCTATAACCCAGTCCTCTGGGGATCTCATCTTGAATGTGGACT
CAGACACCACGATCGCCCCCGACGTCGTCTCGAAGCTTGCCCACAAAATGCGGGATCCAGCAGTCGGTGC
GGCGATGGGCCAAATGAAAGCCAGTAACCAGGCGGACACCTGGCTAACTCGTTTGATTGATATGGAGTAC
TGGATTGCCTGCAACGAGGAGCGCGCGGCACAAGCTCGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCAT
GTGCGATGTACCGTCGGTCTGCTATGCTTTCGCTGCTCGATCAGTACGAGACGCAGCTTTATCGCGGCAA
GCCGAGTGACTTCGGCGAAGATCGCCATTTGACGATTCTCATGCTGAGCGCAGGCTTTCGAACTGAGTAT
GTTCCGAGCGCCATCGCGGCGACAGTCGTTCCAAACACGATGGGTGTTTATCTGCGTCAACAACTACGGT
GGGCACGCAGCACCTTTCGGGATACTTTGCTTGCACTTCCCCTACTGCCTGGTCTCGATCGGTATCTCAC
GCTGGACGCAATCGGGCAAAATGTCGGCCTTCTACTTCTTGCGCTGTCGGTATTGACAGGAATTGGCCAG
TTTGCGCTGACCGCCACAGTACCCTGGTGGACGATCCTGGTCATTGGATCCATGACTCTTGTACGGTGCA
GCGTGGCTGCCTATCGCGCCCGCGAACTCAGGTTTCTGGGTTTTGCACTCCACACGCTCCTGAACATCTT
TCTCTTAATTCCCTTGAAAGCCTATGCCCTTTGTACCCTATCCAATAGCG

>FR850991.1 *Rhizobium leguminosarum* partial nodC sequence for nodulation protein C, strain R45914
TTCGGTGCCGTGATGTGTTGCTGTGGACCCTGTGCCATGTACCGCCGGTCCGCCTTCGTCTTGCTTCTTG
ATCAATACGAGACGCAGCTTTATCGGGGAAAGCCGAGTGACTTTGGCGAGGACCGTCATCTGACCATCCT
AATGCTAAGCGCCGGCTTCCGCACCGAATATGTCCCAAGCGCCATCGCCGCAACCGTTGTTCCTGACGGT
TTGGCAGCCTATCTGCGCCAGCAACTGCGTTGGGCACGGAGTACATTCCGAGATACAATGCTTGGGCTTC
ACCTCCTCCGCGGCATGAACTGGTATCTGACTTTGGACGTCGTCGGGCAGAATGCCGGCCCTCTTCTGCT
CGCATTGTCCRTGCTGGCGGGTCTTGCACAGTTCGCCCTGACGGGTTCAGTGCCCTGGTGGACGATAGGA
ACGATTGGATCATTGACACTGATACGATGCGGCGTGGCTGCTTATCGTGCCAAGCAACTTAGGTTTCTCG
GCTTCTCGCTGCACACGCTCGTGAA

>FR850944.1 *Rhizobium leguminosarum* partial nodC sequence for nodulation protein C, strain R45915
TGTTGCTGTGGACCCTGTGCCATGTACCGCCGGTCCGCCTTCGTCTTGCTTCTTGATCAATACGAGACGC
AGCTTTATCGGGGAAAGCCGAGTGACTTTGGCGAGGACCGTCATCTGACCATCCTAATGCTAAGCGCCGG
CTTCCGCACCGAATATGTCCCAAGCGCCATCGCCGCAACCGTTGTTCCTGACGGTTTGGCAGCCTATCTG
CGCCAGCAACTGCGTTGGGCACGGAGTACATTCCGAGATACAATGCTTGGGCTTCACCTCCTCCGCGGCA
TGAACTGGTATCTGACTTTGGACGTCGTCGGGCAGAATGCCGGCCCTCTTCTGCTCGCATTGTCCRTGCT
GGCGGGTCTTGCACAGTTCGCCCTGACGGGTTCAGTGCCCTGGTGGACGATAGGAACGATTGGATCATTG
ACACTGATACGATGCGGCGTGGCTGCTTATCGTGCCAAGCAACTTAGGTTTCTCGGCTTCTCSCTGCACA
CGCTCGTGAA

>KJ923047.1 *Rhizobium leguminosarum* bv. *viciae* strain K NodC (nodC) gene, partial cds
CGCTTCGGTGCAGTTATGTGTTGCTGCGGCCCATGTGCGATGTACCGTCGGTCTGCTATGCTTTCGCTGC
TCGATCAGTACGAGACGCAGCTTTATCGCGGCAAGCCGAGTGACTTCGGTGAAGATCGCCATTTGACGAT
TCTCATGCTGAGCGCAGGCTTTCGAACTGAGTATGTTCCGAGCGCCATCGCGGCGACAGTCGTTCCTGAC
ACAATGGGTGTTTATTTGCGTCAACAACTACGGTGGGCACGCAGCACCTTTCGGGATACTTTGCTTGCGC
TCCCCATACTGCCTGGCCTCGATCGGTATCTCACTCTGGACGTAATCGGGCAAAATGGCGGCCTTCTGCT
TCTTGCCCTGTCGGTATTGACGGGTATTGGCCAGTTTGCGCTGACCGCCACAGTACCATGGTGGACGATC
CTGGTGATCGGATCCATGACTCTTGTACGATGCAGCGTGGCTGCCTATCGCGCCCGCGAACTTAGGTTTC
TGGGTTTTGCTCTCCACACGCTCGTGAACATCTTTCTCTTAATTCCCTTGAAGGCCTATGCCCTTTGTAC
CTTGTC

>FR851010.1 *Rhizobium leguminosarum* partial nodC sequence for nodulation protein C, strain R46082
GCCTGCAATGAGGAGCGCGCCGCGCAGGGACGCTTCGGTGCCGTGATGTGTTGCTGTGGACCCTGTGCCA
TGTACCGCCGGTCCGCCTTCGTCTTGCTTCTTGATCAATACGAGACGCAGCTTTATCGGGGAAAGCCGAG
TGACTTTGGCGAGGACCGTCATCTGACCATTCTAATGCTAAGCGCCGGCTTCCGTACCGAATATGTCCCA
AGCGCCATCGCCGCAACCGTCGTTCCTGACGGTTTGGCTGCCTATCTGCGCCAGCAACTGCGTTGGGCAC
GGAGTACATTCCGAGATACCCTGCTCGGACTTCACCTCCTCCGTGGCATGAACTGGTATCTGACTTTGGA
CGTCGTCGGGCAGAATGCCGGCCCTCTTCTGCTCGCATTCTCCGTACTGGCGGGTCTAGCACAGTTCGCC
CTGACCGGTTCACTGCCCTGGTGGACGATAGGAACGATTGGATCGTTGACACTGATACGATGCGGCGTAG
CTGCTTATCGTGCCAGGCAGCTTAGATTTCTTGGTTTCTCGCTGCACACGCTCGTGAACGTCTTTCTATT
ACTGCCGGTGAAGGCTTATGCCCTCTGCACGTTGTC

>KT869776.1 *Neorhizobium galegae* strain G226 N-acetylglucosaminyltransferase (nodC) gene, partial cds
GTGAGGCAATCTTGCCTGTTCACGGGCTTTACACCTCGGATCCGAGGTTCGAATTCCTGCTGCTGCCGAA
GAATGTCGGAAAGCGCAAAGCACAGATCGCTGCAATAGAACGGTCATGTGGAGACTTGATCCTAAACGTT
GATTCGGACACTTCTATCGCCTCGGACGTCGTGACTCTGCTCGTCGAAAAAATGCGCGACTCTGACGTTG
GCGCCGCTATGGGCCAATTAAAAGCGAGCAATCGCGACCAAAACCTGCTGACGCGCTTGATTGACATGGA
ATACTGGTTAGCCTGTAACGACGAACGCGCCGCCCAGGCTCGATTTGGCGCGGTTATGTGTTGCTGCGGC
CCGTGCGCCATGTACCGCCGGTCCGCACTTCTCTTGCTTCTTGATCAGTATCAAACGCAACTTTATCGAG
GCAAGCCAAGTGACTTTGGCGAGGATCGCCATTTAACGATCCTCATGCTAAGTGCAGGATTCAGAACCGA
ATACGTTCCCGAGGCAATAGCGAAAACTGTCGTACCAGACCGCATAGGGTCCTACCTACGTCAGCAACTA
CGTTGGGCACGCAGCACCTTCCGTGACACCTTACTCGCTTTGCCTCTCCTGCCTAGCCACAATCGCTTTC
TGACATTGGATGCCATTCACCAGAACATCGGACCATTGCTCTTGGCGGTGTCTTCAGCCACGGGCATAAT
CCAGTTCGTCCTAACAGCAACGATGCCGGGATGGACGATCATCATCATTGCGTCGATGACAATGGTGCGA
TGTTCAGTCGCTGCCTACCGCGCGC

>KM361899.2 *Neorhizobium galegae* strain MQ-5s N-acetylglucosaminyltransferase (nodC) gene, partial cds
CCGCAACCCTATGTTCAATATCATTCTGCTTCCCGACAATGTCGGGAAGCGCAAGGCGCAGATCGCGGCG
ATACGCCGCTCATATGGAGATTTGGTGCTTAACGTTGACTCTGACACGATACTCGCGTCCGACGTCATCG
CAAAGCTTGCGTTGAAGATGCGCAATCCAACGATCGGCGCGGCCATGGGCCAGTTGACGGCTAGCAACCG
GAGCGACACCTGGCTGACCCGGATGATCGATATGGAGTACTGGCTGGCTTGCAACGAGGAGCGCGCGGCA
CAGGCTCGCTTCGGTGCCGTGATGTGCTGCTGCGGCCCATGTGCCATGTATCGCCGATCCGCGCTTCTTT
TGCTGCTGGATAAGTACGAATCGCAATTTTTTCGGGGAAAGCCAAGCGACTTCGGCGAGGATCGCCATCT

CACGATCCTCATGCTCAAAGCAGGCCTTCTAACCGAGTACGTCCCGGACGCCATCGCGTCGACAGTTGTT
CCTGACAGGCTAGGGCCATATCTGCGCCAACAACTGCGCTGGGCACGTAGCACTTTTCGGGACACGTTGC
TTGCGTTGCGCCTGCTGCCCGGCCTCGATCGCTTTCTCACATTGGACGTGATCGGACAGAATCTCGGTCC
GCTACTCCTCGCGCTATCGGTGCTAACCGGGCTCGCGCAG

>KT869794.1 *Neorhizobium galegae* strain G321 N-acetylglucosaminyltransferase (nodC) gene, partial cds
GTGAGGCAATCTTGCCTGTTCACGGGCTTTACACCTCGGATCCGAGGTTCGAATTCCTGCTGCTGCCGAA
GAATGTCGGAAAGCGCAAAGCACAGATCGCTGCAATAGAACGGTCATGTGGAGACTTGATCCTAAACGTT
GATTCGGACACTTCTATCGCCTCGGACGTCGTGACTCTGCTCGTCGAAAAAATGCGCGACTCTGACGTTG
GCGCCGCTATGGGCCAATTAAAAGCGAGCAATCGCGACCAAAACCTGCTGACGCGCTTGATTGACATGGA
ATACTGGTTAGCCTGTAACGACGAACGCGCCGCCCAGGCTCGATTTGGCGCGGTTATGTGTTGCTGCGGC
CCGTGCGCCATGTACCGCCGGTCCGCACTTCTCTTGCTTCTTGATCAGTATCAAACGCAACTTTATCGAG
GCAAGCCAAGTGACTTTGGCGAGGATCGCCATTTAACGATCCTCATGCTAAGTGCAGGATTCAGAACCGA
ATACGTTCCCGAGGCAATAGCGAAAACTGTCGTACCAGACCGCATAGGGTCCTACCTACGTCAGCAACTA
CGTTGGGCACGCAGCACCTTCCGTGACACCTTACTCGCTTTGCCTCTCCTGCCTAGCCACAATCGCTTTC
TGACATTGGATGCCATTCACCAGAACATCGGACCATTGCTCTTGGCGGTGTCTTCAGCCACGGGCATAAT
CCAGTTCGTCCTAACAGCAACGATGCCGGGATGGACGATCATCATCATTGCGTCGATGACAATGGTGCGA
TGTTCAGTCGCTGCCTACCGCGCGC

>FJ499312.1 *Sinorhizobium meliloti* strain RTF15 NodC (nodC) gene, partial cds
ATTCGCGCGATCCGAGGTTCAGCTTCATTCTGCTCCCGGAGAACGTCGGAAAGCGGAAAGCGCAGATTGC
CGCGATAGGCCAATCCTCCGGAGATTTGGTGCTGAATGTCGACTCGGACAGCACGATCGCTTTCGATGTG
GTCTCCAAGCTTGCCTCGAAGATGCGAGATCCAGAGGTCGGTGCGGTTATGGGTCAACTCACGGCTAGCA
ATTCGGGTGACACTTGGCTGACGAAATTGATCGATATGGAGTATTGGCTTGCCTGCAACGAAGAACGCGC
GGCACAGTCTCGCTTCGGTGCTGTTATGTGCTGCTGCGGCCCTTGTGCTATGTACCGCCGGTCGGCGCTC
GCTTCGCTGCTTGACCAGTACGAAACGCAACTGTTTCGCGGTAAGCTAAGCGACTTCGGTGAGGACCGCC
ATCTGACGATCCTCATGTTGAAGGCAGGGTTTCGAACTGAGTATGTTCCAAACGCCATAGTGGCAACCGT
TGTCCCGGATACGCTGAAATCGTATCTGCGCCAACAACTGCGTTGGGCACGCAGCACGTTCCGCGACACA
TTTCTAGCGCTCCCTCTGTTGCGCGGCCTCAACCCTTTTCTCACATTTGATGTAGTCGGGCAGAATATCG
GGCCACTGTTGCTTGCTCTGTCGGTGGTGACGGGACTCGCGCATTTCATAACGACCGCCACAGTGCCATG
GTGGACAATTTTGATTATTGCGTCCATGACCATCATACGCTGCAGCGTCGTAGCATTGCATGCTCGCCAA
CTTAGATTTCTTGGCTTCGTTCTGCACACACCCATCAACCTCTTTCTCTTNCTTCCGTTGAAAGCTTATG
CGTTGGA

>KF749155.1 *Sinorhizobium meliloti* strain SMX12-1 N-acetylglucosaminyl transferase (nodC) gene, partial cds
AAGAACGCGCGGCACAGTCTCGCTTCGGTGCTGTTATGTGTTGCTGCGGCCCTTGTGCTATGTACCGTCG
GTCGGCGCTCGCTTCGCTGCTTGACCAGTACGAAACGCAACTGTTTCGCGGTAAGCCAAGCGACTTCGGT
GAGGACCGCCATCTGACGATTCTCATGTTGAAGGCAGGCTTTCGAACTGAGTACGTTCCAGACGCCATAG
TGGCAACCGTCGTCCCGGATACGCTGAAACCATATCTGCGCCAACAACTGCGTTGGGCACGCAGCACGTT
CCGTGACACGTTTCTAGCGCTCCCTCTGTTGCGCGGCCTCAGCCCTTTTCTCGCATTTGACGCGGTCGGA
CAGAATATCGGGCAACTGTTGCTCGCCCTGTCGGTGGTGACGGGTCTTGCGCATCTCATAATGACCGCCA
CAGTGCCATGGTGGACAATTTTGATTATTGCGTGCATGACCATTATACGCTGCAGCGTCGTAGCATGGCA
TGCTCGCCAACTTAGATTTCTTGGCTTCGTTCTGCACACACCCATCAACCTCTTTCTCATACTTCCGCTG
AAAGCTTATGCGTTGTGTACATTGTCCAACAGCGACTGGCT

>FR851116.1 *Sinorhizobium meliloti* partial nodC sequence for nodulation protein C, strain R46309
GCCTGCAACGAAGAACGCGCGGCACAGTCTCGCTTCGGTGCTGTTATGTGCTGCTGCGGCCCTTGTGCTA
TGTACCGCCGGTCGGCGCTCGCTTCGCTGCTTGACCAGTACGAAACGCAACTGTTTCGCGGTAAGCTAAG
CGACTTCGGTGAGGACCGCCATCTGACGATCCTCATGTTGAAGGCAGGGTTTCGAACTGAGTATGTTCCA
AACGCCATAGTGGCAACCGTTGTCCCGGATACGCTGAAATCGTATCTGCGCCAACAACTGCGTTGGGCAC
GCAGCACGTTCCGCGACACATTTCTAGCGCTCCCTCTGTTGCGCGGCCTCAACCCTTTTCTCACATTTGA
TGTAGTCGGGCAGAATATCGGGCCACTGTTGCTTGCTCTGTCGGTGGTGACGGGACTCGCGCATTTCATA
ACGACCGCCACAGTGCCATGGTGGACAATTTTGATTATTGCGTCCATGACCATCATACGCTGCAGCGTCG
TAGCATTGCATGCTCGCCAACTTAGATTTCTTGGCTTCGTTCTGCACACACCCATCAACCTCTTTCTCTT
ACTTCCGTTGAAAGCTTATGCGTTGTGTACATTGTC

>KP765354.1 *Sinorhizobium meliloti* strain ORT16 N-acetylglucosaminyltransferase (nodC) gene, partial cds
CTCACGGCTAGCAATCCGGGTGACACTTGGCTGACGAAATTGATCGATATGGAGTATTGGCTTGCCTGCA
ACGAAGAACGCGCGGCACAGTCTCGCTTCGGTGCTGTTATGTGCTGCTGCGGCCCTTGTGCTATGTACCG
CCGGTCGGCGCTCGCTTCGCTGCTTGACCAGTACGAAACGCAACTGTTTCGCGGTAAGCTAAGCGACTTC
GGTGAGGACCGCCATCTGACGATCCTCATGTTGAAGGCAGGGTTTCGAACTGAGTATGTTCCAAACGCCA
TAGTGGCAACCGTTGTCCCGGATACGCTGAAATCGTATCTGCGCCAACAACTGCGTTGGGCACGCAGCAC
GTTCCGCGACACATTTCTAGCGCTCCCTCTGTTGCGCGGCCTCAACCCTTTTCTCACATTTGATGTAGTC
GGGCAGAATATCGGGCCACTGTTGCTTGCTCTGTCGGTGGTGACGGGACTCGCGCATTTCATAACGACCG
CCACAGTGCCATGGTGGACAATTTTGATTATTGCGTCCATGACCATCATACGCTGCGGCGTCTTAGCATT
GCATGCTCGCCAACTTAGATTTCTTGGCTTCGTTCTGCACACACCCATCAACCTC

>KP765350.1 *Sinorhizobium meliloti* strain SF3.10 N-acetylglucosaminyltransferase (nodC) gene, partial cds
CGCCTTCTATTCGCGCGATCCGAGGTTCAGCTTCATTCTGCTCCCGGAGAACGTCGGAAAGCGGAAAGCG
CAGATTGCCGCGATAGGCCAATCCTCCGGAGATTTGGTGCTGAATGTCGACTCGGACAGCACGATCGCTT
TCGATGTGGTCTCCAAGCTTGCCTCGAAGATGCGAGATCCAGAGGTCGGTGCGGTTATGGGTCAACTCAC
GGCTAGCAATCCGGGTGACACTTGGCTGACGAAATTGATCGATATGGAGTATTGGCTTGCCTGCAACGAA
GAACGCGCGGCACAGTCTCGCTTCGGTGCTGTTATGTGCTGCTGCGGCCCTTGTGCTATGTACCGCCGGT
CGGCGCTCGCTTCGCTGCTTGACCAGTACGAAACGCAACTGTTTCGCGGTAAGCTAAGCGACTTCGGTGA
GGACCGCCATCTGACGATCCTCATGTTGAAGGCAGGGTTTCGAACTGAGTATGTTCCAAACGCCATAGTG
GCAACCGTTGTCCCGGATACGCTGAAATCGTATCTGCGCCAACAACTGCGTTGGGCACGCAGCACGTTCC
GCGACACATTTCTAGCGCTCCCTCTGTTGCGCGGCCTCAACCCTTTTCTCACATTTGATGTAGTCGGGCA
GAATATCGGGCCACTGTTGCTTGCTCTGTCGGTGGTGACGGGACTCGCGCATTTCATAACGACCGCCACA
GTGCCATGGTGGACAA

Table A 8: Taxonomic classification of the non-*Rhizobium* genera from *16S rRNA* gene sequencing.

Phylum	Class	Order	Family	Genus
Acidobacteria	Acidobacteriia	Acidobacteriales	Acidobacteriaceae	Undefined
			Microbacteriaceae	Undefined
			Propionibacteriaceae	*Propionibacterium*
			Streptomycetaceae	*Streptomyces*
Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	*Bacteroides*
	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	*Flavobacterium*
	Saprospirae	Saprospirales	Chitinophagaceae	Undefined
	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Undefined
Firmicutes	Bacilli	Bacillales	Bacillaceae	*Anaerobacillus*
				Bacillus
				Undefined
			Planococcaceae	*Lysinibacillus*
				Sporosarcina
Gemmatimonadetes	Gemmatimonadetes	N1423WL	Undefined	Undefined
Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	*Asticcacaulis*
				Caulobacter
				Mycoplana
				Phenylobacterium
		Kiloniellales	Kiloniellaceae	*Thalassospira*
		Rhizobiales	Bradyrhizobiaceae	*Bradyrhizobium*
			Brucellaceae	*Ochrobactrum*
			Hyphomicrobiaceae	*Devosia*
			Rhizobiaceae	*Agrobacterium*
				Sinorhizobium
				Undefined
		Rhodospirillales	Rhodospirillaceae	*Inquilinus*
		Sphingomonadales	Sphingomonadaceae	*Novosphingobium*
				Sphingomonas
	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Burkholderia*
			Comamonadaceae	*Acidovorax*
				Delftia
				Rhodoferax
			Oxalobacteraceae	*Janthinobacterium*
				Undefined
	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Erwinia*
				Rahnella
				Salmonella
		Pseudomonadales	Pseudomonadaceae	*Pseudomonas*
		Xanthomonadales	Xanthomonadaceae	*Rhodanobacter*
				Stenotrophomonas
				Undefined
Tenericutes	Mollicutes	Anaeroplasmatales	Anaeroplasmataceae	*Asteroleplasma*
TM7	TM7-1	Undefined	Undefined	Undefined
	TM7-3	Undefined	Undefined	Undefined

Appendix 3: Supplementary data

Chapter 2 supplementary results

Table A 9: Summary of the means of *16S rRNA* OTUs and reads across the pH groups for the two clover species. Group A (pH 4.3 – 5.3), Group B (pH 5.5 – 6.1) and Group C (pH 6.5 – 7.5).

	Subterranean Clover			**White Clover**		
pH group	**A**	**B**	**C**	**A**	**B**	**C**
Mean OTUs	225	213	212	196	194	195
Mean Reads	60,083	55,507	57,029	61,137	58,799	63,736

Beta diversity analysis

The Bray-Curtis resemblance data for both *Rhizobium* and non-*Rhizobium* genera was analysed using BEST analysis (Primer v7) against soil physicochemical properties. This analysis showed very low correlations ($\rho < 0.1$) of the soil properties with beta diversity of *Rhizobium* and non-*Rhizobium* genera, which indicated no/minimum influence of soil properties on the beta diversity across genera.

Table A 10: Pearson's correlations of the relative abundance and H' index for genus *Rhizobium* and all 'other' genera, across SC and WC samples with measured soil physicochemical parameters, and their corresponding p values. Values with * are significant at 95%.

	Rhizobium				**Non-*Rhizobium***			
	Relative abundance		H' index		Relative abundance		H' index	
Subterranean clover (SC)	Pearson's r	p value	Pearson's r	p value	Pearson's r	p value	Pearson's r	p value
Olsen Phosphorus	0.0118	0.9110	0.0133	0.9000	0.0226	0.8305	-0.1003	0.3415
Potassium	0.0007	0.9945	-0.1232	0.2419	-0.2116	0.0429*	-0.1946	0.0631
Calcium	0.0077	0.9416	-0.1712	0.1027	-0.1235	0.2408	0.0352	0.7390
Magnesium	0.0748	0.4787	0.1170	0.2667	-0.0988	0.3490	-0.0566	0.5918
Aluminium	0.1035	0.3260	0.2066	0.0481*	0.0283	0.7888	-0.0611	0.5632
Organic Matter	-0.1802	0.0856	0.0139	0.8956	-0.0387	0.7139	-0.0635	0.5477
White clover (WC)								
Olsen Phosphorus	0.2433	0.0104*	0.0508	0.5981	0.0579	0.5482	0.2111	0.0269*
Potassium	0.1942	0.0420*	0.0519	0.5905	-0.0527	0.5848	0.1149	0.2318
Calcium	0.0618	0.5215	0.0917	0.3406	-0.1425	0.1375	-0.0985	0.3060
Magnesium	0.0680	0.4803	0.0447	0.6426	0.0966	0.3152	0.1945	0.0417*
Aluminium	-0.0747	0.4381	-0.0745	0.4390	0.0707	0.4627	-0.0336	0.7271
Organic Matter	0.0335	0.7286	-0.0674	0.4839	-0.0663	0.4916	0.1333	0.1652

Table A 11: Pearson's correlations of the relative abundance and OTUs of nodC sequences across SC and WC samples with measured soil physicochemical parameters, and their corresponding P values. Values with * are significant at 95% confidence intervals.

	Subterranean clover (SC)				**White clover (WC)**			
	Relative abundance		No. of OTUs		Relative abundance		No. of OTUs	
	Pearson's r	P (95% CI)	Pearson's r	P (95% CI)	Pearson's r	P (95% CI)	Pearson's r	P (95% CI)
Olsen Phosphorus	0.0136	0.8975	-0.0194	0.8547	0.0875	0.3632	-0.0393	0.6832
Potassium	0.0884	0.4021	-0.1221	0.2461	0.1206	0.2095	-0.0718	0.4563
Calcium	0.0343	0.7452	-0.0726	0.4916	0.0670	0.4866	0.0769	0.4244
Magnesium	-0.0302	0.7747	-0.1843	0.0786	-0.0264	0.7845	-0.0185	0.8475
Sodium	-0.0625	0.5536	-0.1383	0.1887	-0.0867	0.3676	-0.0504	0.6008
Potentially Available N	-0.1203	0.2532	-0.0352	0.7393	-0.0466	0.6285	0.0958	0.3194
AMN	-0.1271	0.2273	0.0280	0.7914	0.0001	0.9995	0.1228	0.2013
AMN/Total N Ratio	0.0160	0.8794	0.0410	0.6983	-0.0046	0.9616	0.2012	0.0350 *
Aluminium	-0.0090	0.9322	-0.0192	0.8557	-0.0284	0.7682	-0.1252	0.1925
Organic Matter	-0.1964	0.0606	0.0012	0.9906	-0.0216	0.8230	-0.0621	0.5192
Total Carbon	-0.1462	0.1644	0.0385	0.7159	0.0191	0.8427	0.0028	0.9766
Total Nitrogen	0.0086	0.9348	0.0486	0.6452	0.0941	0.3280	0.0027	0.9779
C/N Ratio	-0.0757	0.4733	-0.0142	0.8934	-0.1097	0.2538	-0.0958	0.3195
CEC	-0.1829	0.0810	-0.2091	0.0454 *	0.0291	0.7631	-0.0167	0.8625

Soil properties

Soil Name	pH	Olsen Phosphorus	Potassium	Calcium	Magnesium	Sodium	Potentially Available Nitrogen	Anaerobically Mineralisable N (AMN)	AMN/Total N Ratio	Aluminium (CaCl2 Extractable)	Organic Matter	Total Carbon	Total Nitrogen	C/N Ratio	CEC	Total Base Saturation	Volume Weight	GPS co-ordinates
	pH Units	mg/L	MAF units	MAF units	MAF units	MAF units	kg/ha	µg/g	%	mg/kg	%	%	%		me/100g	%	g/mL	
Windies	4.3	107	15	6	20	8	228	216	3.3	12.9	13.7	8	0.65	12.3	26	38	0.7	S38 50 41.4 E177 08 26.7
EW_326	4.6	37	7	4	13	5	169	179	2.2	17.6	27.3	15.8	0.81	19.5	30	22	0.63	S45 21 28.9 E167 49 06.1
CF_ICI3	4.7	9	6	6	14	13	173	200	2.3	5.4	26.9	17.6	15.6	0.89	35	29	0.58	S41 47 13.6 E171 34 20.5
Crossing	4.7	58	15	6	23	4	116	117	2.4	7.5	9.8	5.7	0.49	11.7	24	41	0.66	S38 52 41.1 E177 09 42.0
SU_47	4.8	59	10	10	32	6	222	179	2.7	4	14.9	8.6	0.66	13	25	51	0.82	S42 30 27.5 E171 35 35.5
CF_IC15	4.9	12	7	4	15	11	152	157	2.2	9.4	21.4	12.4	0.71	17.5	29	24	0.64	S41 47 10.1 E171 34 25.6
G13+BH29	4.9	10	6	4	13	13	249	260	4.1	2.8	17.4	15.8	10.1	0.64	22	33	0.64	S41 47 45.9 E171 33 44.8
Waimak 55	4.9	54	12	11	36	8	206	158	3.4	4.1	11.3	6.6	0.47	14	24	54	0.87	S43 24 10.8 E172 14 48.6
ARA_254	5.0	61	11	5	10	4	104	99	2	7	10.1	5.9	0.49	12.1	19	41	0.7	S38 38 25.1 E176 09 32.7
Kaitawa	5.0	28	15	7	24	9	249	321	3.8	3.7	18.1	10.5	0.84	12.4	35	40	0.52	S38 53 09.2 E177 02 32.3
Pines	5.0	21	9	6	18	6	281	305	3.6	3.3	15.4	8.9	0.85	10.5	25	40	0.61	S38 52 37.4 E177 00 34.7
TO_58	5.0	26	5	2	11	7	133	91	3.7	13	5.3	3.1	0.25	12.3	9	29	0.97	S41 47 21.5 E171 29 24.6
WE_C31	5.0	34	11	5	20	7	247	247	3.6	2.6	13.5	7.8	0.68	11.5	22	37	0.67	S42 34 24.6 E171 34 35.2
Woods	5.0	75	15	8	19	8	250	259	3.1	4.1	18	10.4	0.82	12.7	27	45	0.64	S38 52 24.9 E177 01 30.8
CH_F34	5.1	54	9	16	21	5	178	167	2.3	2.2	17.9	10.4	0.73	14.2	38	53	0.71	S45 31 41.1 E168 07 43.6
SU_55	5.1	42	7	6	25	6	348	360	4.6	4.1	18.3	10.6	0.78	13.6	25	40	0.64	S42 30 54.5 E171 35 08.9
TO_63	5.1	46	5	4	23	7	123	102	3.2	4	6.7	3.9	0.31	12.3	14	40	0.81	S41 46 38.8 E171 29 30.1
Bills	5.2	44	10	6	17	6	247	266	3.5	3.9	16.1	12.2	9.4	0.77	25	39	0.62	S38 51 12.7 E177 08 13.9
CF_G28	5.2	24	5	5	10	12	225	179	4.5	2.5	9.3	5.4	0.4	13.4	14	41	0.84	S41 49 04.9 E171 32 36.3
ER_613	5.2	114	16	13	21	2	194	208	3.1	2.4	15.3	8.9	0.67	13.3	32	61	0.62	S38 33 48.3 E176 14 30.3
MD_R22	5.2	45	10	11	26	10	268	232	3.6	2.9	10.7	6.2	0.64	9.7	27	52	0.77	S43 53 09.5 E171 33 03.6
WE_C22	5.2	23	8	4	15	9	258	265	4	4.3	15.6	13.8	9	0.66	23	30	0.65	S42 34 52.7 E171 33 42.5
ARI_4+5	5.3	84	16	10	35	17	151	155	2.3	2.1	14.7	8.5	0.69	12.4	29	57	0.65	S38 34 36.7 E176 18 59.9
Hills	5.3	94	40	13	35	6	158	124	3	0.8	7.4	4.3	0.42	10.1	25	65	0.85	S39 26 02.9 E176 50 56.6
WE_C39	5.3	40	5	7	9	8	337	306	5.7	2.2	11.1	6.5	0.54	12	21	40	0.73	S42 34 45.8 E171 34 32.1
AR_120	5.5	44	7	5	13	4	226	297	3.7	5.2	17.2	12.4	10	0.81	29	33	0.51	S38 37 42.5 E176 09 07.1
CH_G29	5.6	39	11	14	16	5	201	208	2.7	1.3	18.2	10.6	0.78	13.6	33	60	0.64	S45 28 40.5 E168 06 33.5
DG_D36	5.6	10	6	6	34	3	188	112	3.3	0.8	5.4	9.3	3.1	0.34	13	48	1.12	S45 37 57.9 E167 40 13.9
MD_R28	5.6	37	15	12	25	7	205	161	2.9	1.8	10.7	11.2	6.2	0.55	24	59	0.85	S43 53 25.6 E171 32 56.4
Waimak 39	5.6	14	4	8	14	3	66	49	1.8	1.6	5.5	3.2	0.28	11.7	16	49	0.91	S43 24 49.9 E172 15 10.8
ARA_74	5.7	65	18	6	21	5	202	262	3.3	1.4	17.7	10.3	0.79	13	28	48	0.51	S38 37 49.0 E176 08 10.9
TO_39	5.8	35	6	7	15	6	153	115	2.7	1.1	7	9.5	4.1	0.43	12	63	0.89	S41 47 30.5 E171 30 03.3
ARI_46	6.0	116	30	9	33	5	232	291	2.8	2.4	21.2	11.9	12.3	1.03	31	64	0.53	S38 35 19.8 E176 20 23.6
DG_D16	6.0	28	7	9	13	3	250	213	2.7	1	17.6	12.8	10.2	0.8	19	57	0.78	S45 37 37.4 E167 39 51.6
Edwards 6	6.0	68	30	20	107	96	187	177	2.9	< 0.2	9.6	5.5	0.6	9.2	40	88	0.7	S39 27 22.2 E176 50 42.2
ARI_48	6.1	58	9	12	14	4	155	180	2.9	2.5	13.3	12.4	7.7	0.62	23	78	0.58	S38 35 19.8 E176 20 23.6
CH_F38	6.1	29	9	16	18	5	223	215	3.3	1.3	16.4	14.7	9.5	0.65	33	62	0.69	S45 31 21.9 E168 07 25.4
TO_C6	6.1	21	5	5	9	8	41	25	1.2	1	5.3	14.3	3.1	0.22	7	57	1.1	S41 46 30.1 E171 29 54.7
DG_W25	6.5	34	12	22	17	7	168	162	2.4	0.2	16.7	9.7	0.69	14.1	30	91	0.69	S45 36 01.0 E167 41 08.3
Pump Triangle	6.5	58	20	23	38	8	114	83	2.9	< 0.2	4.7	2.7	0.29	9.5	25	90	0.92	S39 28 38.8 E176 51 10.7
Watchmen 1	6.5	59	27	22	63	16	153	125	2.7	< 0.2	7	4.1	0.46	8.9	31	88	0.82	S39 27 26.1 E176 50 30.7
Waimak 11	6.5	20	13	12	18	6	97	71	1.9	0.3	7	10.7	4.1	0.38	17	72	0.91	S43 25 54.1 E172 15 47.1
ER_38	6.6	60	40	11	27	11	97	94	3	1	8.7	16	5.1	0.32	19	92	0.68	S38 34 48.1 E176 15 05.9
Back Beacons	6.7	33	24	19	193	392	113	86	1.7	< 0.2	7.8	4.6	0.5	9.1	40	95	0.88	S39 27 24.5 E176 51 01.2
SiberiaWatch	7.1	43	20	23	56	23	130	103	2.6	< 0.2	6.2	9	3.6	0.4	27	100	0.84	S39 28 21.5 E176 50 59.5
Pidgeon13	7.5	44	10	29	23	9	168	111	2.7	< 0.2	5.3	7.5	3.1	0.41	25	100	1.01	S39 28 18.5 E176 51 05.4

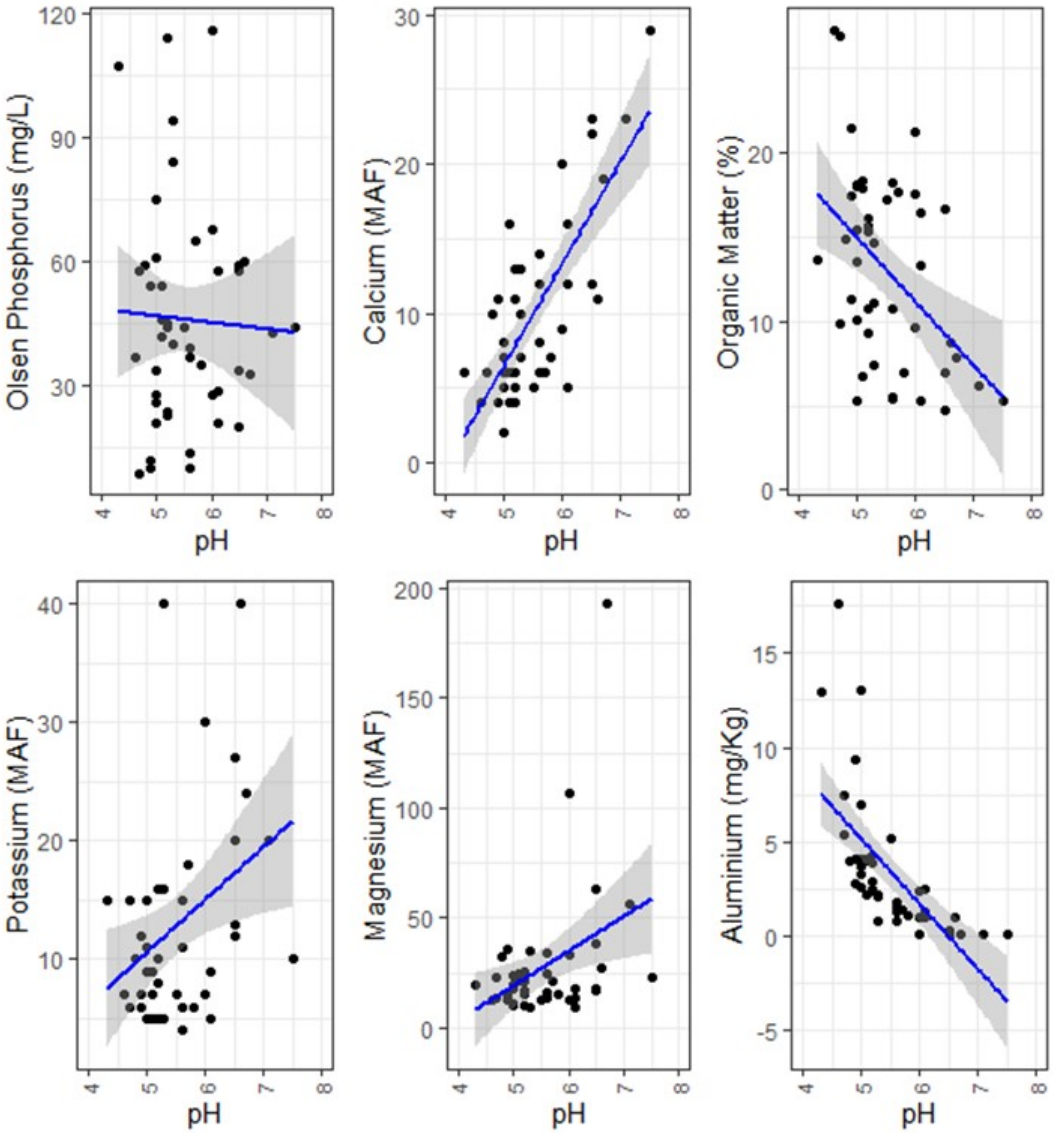

Figure A 1: Correlations of selected soil properties with pH.

Table A 12: Summary of the baiting experiment showing number of samples and nodules for each clover species, and pH groups the soils were split into. SC = Subterranean clover, WC = White clover.

			SC		WC	
Soil	**Soil pH**	**pH group**	**No. of samples**	**No. of nodules**	**No. of samples**	**No. of nodules**
Windies	4.3	Group A	0	0	1	8
EW_326	4.6	Group A	2	33	1	5
Crossing	4.7	Group A	0	0	3	110
SU_47	4.8	Group A	3	76	3	44
Waimak 55	4.9	Group A	1	38	2	20
CF_IC15	4.9	Group A	2	39	3	49
G13+BH29	4.9	Group A	2	62	2	32
TO_58	5.0	Group A	3	88	3	103
Woods	5.0	Group A	1	10	2	28
Kaitawa	5.0	Group A	0	0	3	75
Pines	5.0	Group A	3	55	3	85
WE_C31	5.0	Group A	2	31	3	75
ARA_254	5.0	Group A	3	66	3	40
CH_F34	5.1	Group A	2	38	0	0
SU_55	5.1	Group A	0	0	3	80
TO_63	5.1	Group A	3	104	3	103
CF_G28	5.2	Group A	3	84	3	106
ER_613	5.2	Group A	3	85	3	95
MD_R22	5.2	Group A	3	64	3	81
ARI_4+5	5.3	Group A	1	30	3	80
WE_C39	5.3	Group A	3	112	3	95
Hills	5.3	Group A	3	67	3	76
AR_120	5.5	Group B	0	0	1	8
Waimak 39	5.6	Group B	3	77	3	64
CH_G29	5.6	Group B	3	112	3	105
MD_R28	5.6	Group B	2	21	1	12
DG_D36	5.6	Group B	1	2	3	35
ARA_74	5.7	Group B	3	96	3	70
TO_39	5.8	Group B	3	56	3	103
Edwards 6	6.0	Group B	3	100	3	101
ARI_46	6.0	Group B	3	95	3	95
TO_C6	6.1	Group B	3	102	3	78
CH_F38	6.1	Group B	2	40	3	92
ARI_48	6.1	Group B	0	0	2	41
Watchmen 1	6.5	Group C	3	105	3	105
Pump Triangle	6.5	Group C	3	108	3	103
DG_W25	6.5	Group C	3	73	3	65
Waimak 11	6.5	Group C	3	101	3	60
ER_38	6.6	Group C	2	52	3	74
Back Beacons	6.7	Group C	3	64	2	14
SiberiaWatch	7.1	Group C	3	94	3	102
Pidgeon13	7.5	Group C	3	93	3	109

Table A 13 summarises the results for SC and WC according to the pH groups. The number of samples per pH group differ and hence there are more nodules in the low pH group compared with medium and high groups for both SC and WC. However, the number of nodules/sample are similar in all three pH groups.

Table A 13: Summary of the number of nodules and samples and the nodules/sample for each clover species based on the pH groups.

	Subterranean Clover			**White Clover**		
	No. of nodules	No. of samples	Nodules/ samples	No. of nodules	No. of samples	Nodules/ samples
Group A (pH < 5.5)	1082	43	25	1390	56	25
Group B (pH 5.5–6.1)	701	26	27	804	31	26
Group C (pH > 6.1)	690	23	30	632	23	27

Chapter 3 supplementary results

Pilots 1 and 2 results

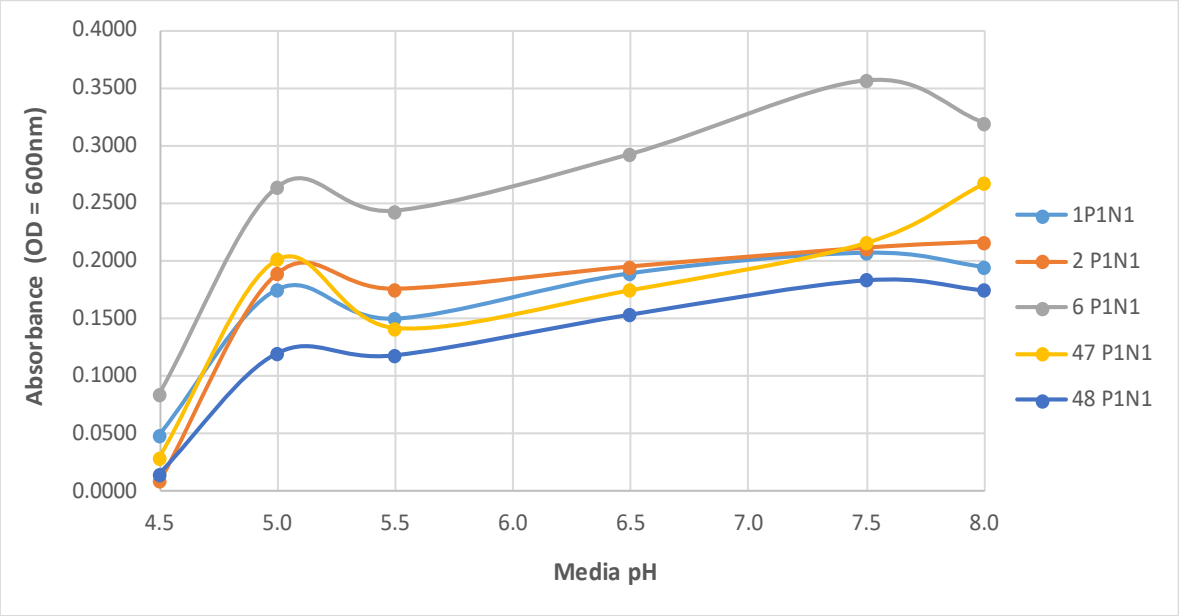

Figure A 2: Pilot 1 results, showing growth curves for the five white clover strains in the corresponding pH-amended media.

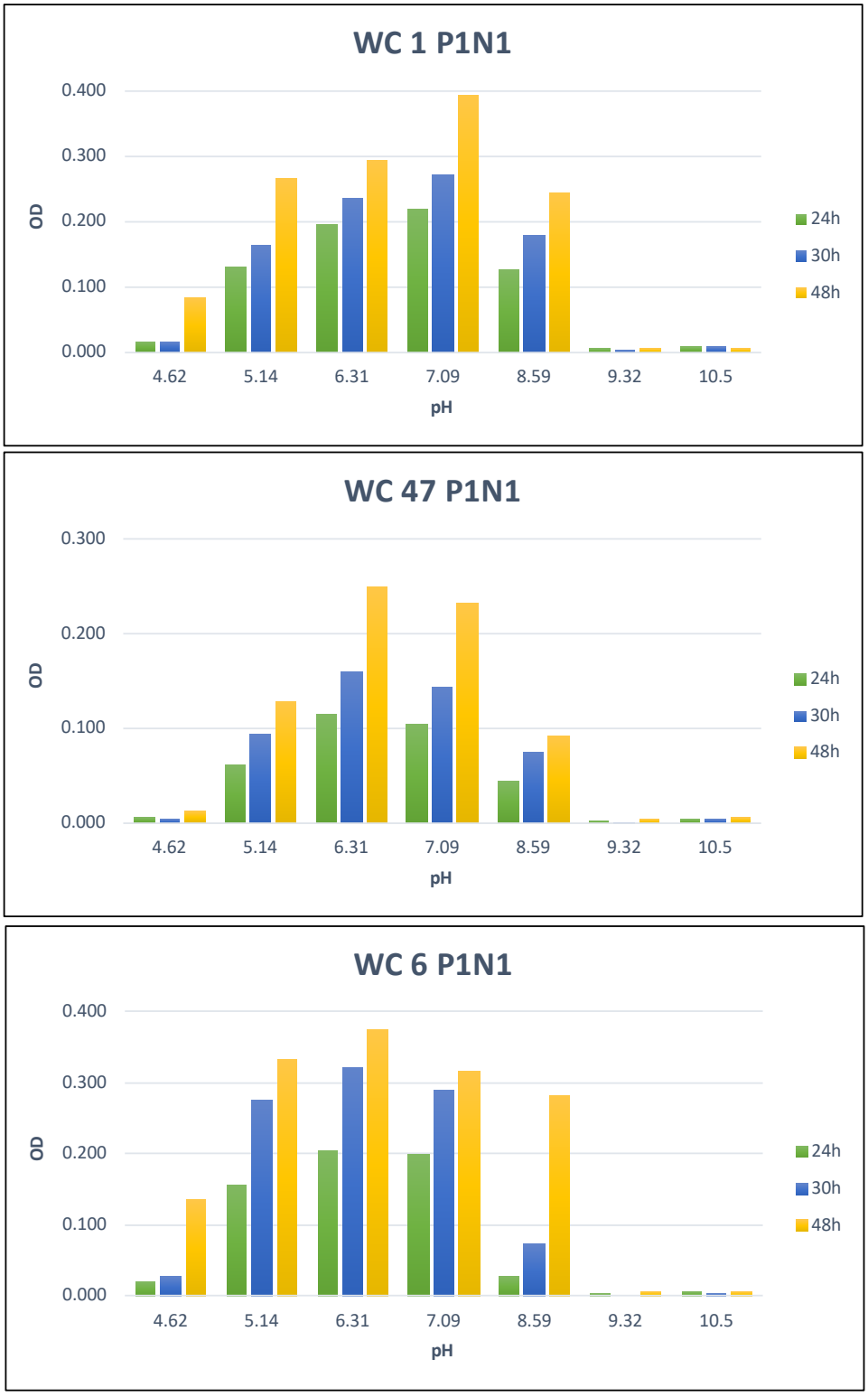

Figure A 3: Pilot 2 results, showing growth (optical density, OD) patterns for three white clover (WC) strains in the corresponding pH-amended media across three time-points.

Bioassay statistical analysis

Table A 14: T-test results for all six batches of WC strains, with Batch 1 as the intercept, i.e. the reference point. Data is for OD measurements taken only at 48h.

	Diff. from Batch 1	Estimated mean	Std. error	t value	Pr (>|t|)
Batch 1	0	0.1974	NA	NA	NA
Batch 2	-0.0599	0.1375	0.0402	-1.491	0.1363
Batch 3	-0.0833	0.1141	0.0568	-1.465	0.1433
Batch 4	-0.0674	0.1300	0.0402	-1.678	0.0938
Batch 5	-0.0059	0.1915	0.0402	-0.149	0.8815
Batch 6	0.0174	0.2148	0.0402	0.433	0.6649

Table A 15: T-test result from comparing Batch 6 with Batches 1–5, with the mean of Batches 1-5 as the intercept. Data is for OD measurements taken only at 48h.

	Diff. from Batch 1-5	Estimated mean	Std. error	t value	Pr (>|t|)
Batch 1-5	0	0.1641	NA	NA	NA
Batch 6	0.0508	0.2149	0.0321	1.578	0.1148

Table A 16: T-test results for batches 2–8 of SC strains, with Batch 2 as the intercept, i.e. the reference point. Data is for ODs averaged across the three time points. P values with * are significantly different from Batch 2 at 95% CI.

	Diff. from Batch 2	Estimated mean	Std. error	t value	Pr (>|t|)
Batch 2	0	0.0961	NA	NA	NA
Batch 3	0.0018	0.0979	0.0082	0.217	0.8281
Batch 4	0.0293	0.1254	0.0082	3.586	0.0004 **
Batch 5	-0.0099	0.0862	0.0082	-1.207	0.2278
Batch 6	-0.0193	0.0768	0.0082	-2.365	0.0183 *
Batch 7	0.0007	0.0968	0.0082	0.089	0.9292
Batch 8	-0.0018	0.0943	0.0082	-0.226	0.8210

Media optimum pH and soil properties

Positive correlations of S ($r = 0.851$, $P < 0.001$) and Na ($r = 0.699$, $P = 0.011$) were present with media optimum pH for WC strains. Whereas, both Ca ($r = 0.8756$, SC; $r = 0.6096$, WC) and Mg ($r = 0.7176$, SC; $r = 0.6811$, WC) had significant positive correlations with media optimum pH for strains of SC and WC. For WC strains, the following soil properties had significant negative correlations with media optimum pH: AMN ($r = -0.599$, $P = 0.039$), OM ($r = -0.664$, $P = 0.018$), total C ($r = -0.664$, $P = 0.019$) and total N ($r = -0.595$, $P = 0.041$). However, for SC strains only the C/N ratio had significant negative correlation with media optimum pH ($r = -0.651$, $P = 0.022$).

Table A 17: Pearson's correlations of media optimum pH for strains of SC and WC with measured soil physicochemical parameters, and their corresponding p values. Values with * are significant at 95% confidence intervals.

	SC		WC	
	Media Optimum pH (40,50h)		Media Optimum pH (38,48h)	
	Pearson's coefficient (r)	P value (95% CI)	Pearson's coefficient (r)	P value (95% CI)
Soil pH	0.6677	0.018 *	0.5674	0.054
Olsen Phosphorus	0.3347	0.288	0.4638	0.129
Sulphate Sulphur	0.4387	0.154	0.8509	<0.001 *
Potassium	0.1762	0.584	0.4337	0.159
Calcium	0.8756	<0.001 *	0.6096	0.035 *
Magnesium	0.7176	0.009 *	0.6811	0.015 *
Sodium	0.469	0.124	0.6996	0.011 *
Potentially Available N	0.1924	0.549	-0.5118	0.089
AMN	0.0761	0.814	-0.5995	0.039 *
AMN/Total N Ratio	0.125	0.699	-0.2962	0.350
Aluminium	-0.4143	0.181	-0.4341	0.159
Organic Matter	-0.1154	0.721	-0.6644	0.018 *
C/N Ratio	-0.6514	0.022 *	-0.4455	0.147
Total Carbon	-0.1161	0.719	-0.664	0.019 *
Total Nitrogen	0.0937	0.772	-0.5953	0.041 *
CEC	0.5735	0.051	0.1274	0.693

Symbiotic potential of selected SC and WC strains

Table A 18: ANOVA results showing Tukey's 95% confidence intervals for shoot dry weights of subterranean and white clover plants inoculated with their respective strains.

SC strain	Mean shoot dry weight (mg/plant)	Tukey's test	WC strain	Mean shoot dry weight (mg/plant)	Tukey's test
42P1N1	25.12	a	4P1N1	3.92	a
84P2N2	30.16	a	9P1N1	7.64	ab
40P1N1	37.00	ab	46P2N1	7.65	ab
38P1N1	38.23	ab	74P3N2	10.57	ab
78P2N1	38.98	ab	41P3N2	10.97	ab
6P2N2	39.70	ab	1P1N1	11.34	ab
6P1N1	44.01	ab	81P2N2	20.02	b
79P3N2	52.55	ab	TA1	20.79	b
83P1N1	57.28	ab	47P2N1	20.93	b
WSM1325	70.25	b	79P2N2	21.39	b

Chapter 4 supplementary results

Omnilog carbon sources

Table A 19: Carbon sources which had AUCs higher than the set threshold (PM01 = 71 C-sources; PM02 = 50 C-sources).

C-Sources (PM01)	C-type	C-Sources (PM02)	C-type
L-Arabinose	Carbohydrate	Dextrin	Polymer
N-Acetyl-D-GlucosAmine	Carbohydrate	Glycogen	Polymer
Succinic acid	Carboxylic acid	Inulin	Polymer
D-Galactose	Carbohydrate	Laminarin	Polymer
L-Aspartic acid	Amino acid	Mannan	Polymer
L-Proline	Amino acid	Pectin	Polymer
D-Trehalose	Carbohydrate	N-Acetyl-D-GalactosAmine	Carbohydrate
D-Mannose	Carbohydrate	β-D-Allose	Carbohydrate
Dulcitol	Carbohydrate	Amygdalin	Carbohydrate
D-Sorbitol	Carbohydrate	D-Arabinose	Carbohydrate
Glycerol	Carbohydrate	D-Arabitol	Carbohydrate
L-Fucose	Carbohydrate	L-Arabitol	Carbohydrate
D-Gluconic acid	Carboxylic acid	Arbutin	Carbohydrate
D,L-α-Glycerol Phosphate	Carbohydrate	2-Deoxy-D-Ribose	Carbohydrate
D-Xylose	Carbohydrate	i-Erythritol	Carbohydrate
L-Lactic acid	Carboxylic acid	D-Fucose	Carbohydrate
Formic acid	Carboxylic acid	3-0-β-D-Galactopyranosyl-D-Arabinose	Carbohydrate
D-Mannitol	Carbohydrate	Gentiobiose	Carbohydrate
L-Glutamic acid	Amino acid	D-Lactitol	Carbohydrate
D-Glucose-6-Phosphate	Carbohydrate	Maltitol	Carbohydrate
D-Galactonic acid-γ-Lactone	Carboxylic acid	α-Methyl-D-Glucoside	Carbohydrate
D,L-Malic acid	Carboxylic acid	β-Methyl-D-Galactoside	Carbohydrate
D-Ribose	Carbohydrate	α-Methyl-D-Mannoside	Carbohydrate
Tween 20	Fatty acid	β-Methyl-D-Xyloside	Carbohydrate
L-Rhamnose	Carbohydrate	Palatinose	Carbohydrate
D-Fructose	Carbohydrate	D-Raffinose	Carbohydrate
Acetic acid	Carboxylic acid	Salicin	Carbohydrate
α-D-Glucose	Carbohydrate	L-Sorbose	Carbohydrate
Maltose	Carbohydrate	D-Tagatose	Carbohydrate
D-Melibiose	Carbohydrate	Turanose	Carbohydrate
L-Asparagine	Amino acid	Xylitol	Carbohydrate
D-Glucosaminic acid	Carboxylic acid	g-Amino-N-Butyric acid	Carboxylic acid
Tween 40	Fatty acid	Butyric acid	Carboxylic acid
α-Methyl-D-Galactoside	Carbohydrate	D-GlucosAmine	Carbohydrate
α-D-Lactose	Carbohydrate	β-Hydroxybutyric acid	Carboxylic acid
Lactulose	Carbohydrate	5-Keto-D-Gluconic acid	Carboxylic acid

Sucrose	Carbohydrate	Melibionic acid	Carbohydrate
Uridine	Carbohydrate	Oxalomalic acid	Carboxylic acid
L-GlutAmine	Amino acid	Quinic acid	Carboxylic acid
D-Glucose-1-Phosphate	Carbohydrate	Succinamic acid	Carboxylic acid
D-Fructose-6-Phosphate	Carbohydrate	L-AlaninAmide	Amide
Tween 80	Fatty acid	L-Arginine	Amino acid
β-Methyl-D-Glucoside	Carbohydrate	L-Histidine	Amino acid
Adonitol	Carbohydrate	Hydroxy-L-Proline	Amino acid
Maltotriose	Carbohydrate	L-Lysine	Amino acid
Adenosine	Carbohydrate	L-Ornithine	Amino acid
Gly-Asp	Amino acid	L-Pyroglutamic acid	Amino acid
m-Inositol	Carbohydrate	D,L-Carnitine	Carboxylic acid
Fumaric acid	Carboxylic acid	D,L-OctopAmine	Amine
Bromosuccinic acid	Carboxylic acid	Dihydroxyacetone	Alcohol
Mucic acid	Carboxylic acid		
D-Cellobiose	Carbohydrate		
Inosine	Carbohydrate		
Gly-Glu	Amino acid		
L-Serine	Amino acid		
L-Alanine	Amino acid		
Ala-Gly	Amino acid		
Acetoacetic acid	Carboxylic acid		
N-Acetyl-D-MannosAmine	Carbohydrate		
Mono-Methylsuccinate	Carboxylic acid		
Methylpyruvate	Ester		
D-Malic acid	Carboxylic acid		
L-Malic acid	Carboxylic acid		
Gly-Pro	Amino acid		
p-Hydroxyphenyl Acetic acid	Carboxylic acid		
D-Psicose	Carbohydrate		
L-Lyxose	Carbohydrate		
GlucuronAmide	Amide		
Pyruvic acid	Carboxylic acid		
D-Galacturonic acid	Carboxylic acid		
2-Aminoethanol	Alcohol		

Table A 20: Climate data for each sampling location of the 19 strains. All data expressed as average over the 2014–2015 calendar year.

Strain	Temp max (°C)	Temp min (°C)	Penman evaporation (Kg/m^2)	Rainfall (mm)	Soil moisture (Kg/m^2 over 24 h)	Solar radiation (MJ/ m^2 over 24 h)
1P1N1	16.622	9.027	2.308	6.239	-13.823	13.318
41P1N2	16.450	9.096	2.294	6.056	-14.986	13.496
41P3N2	16.450	9.096	2.294	6.056	-14.986	13.496
43P2N2	14.111	4.366	1.952	2.872	-33.247	12.617
46P2N1	15.734	7.105	2.318	7.083	-7.189	12.133
47P2N1	19.336	9.093	2.904	1.817	-74.810	15.773
4P1N1	17.401	7.267	2.478	2.039	-59.582	15.140
5P1N2	16.450	9.096	2.294	6.056	-14.986	13.496
5P2N1	16.450	9.096	2.294	6.056	-14.986	13.496
5P2N2	16.450	9.096	2.294	6.056	-14.986	13.496
6P1N2	16.450	9.096	2.294	6.056	-14.986	13.496
73P1N2	16.622	9.027	2.308	6.239	-13.823	13.318
74P3N2	17.168	7.339	2.768	3.929	-25.778	13.398
78P2N2	16.450	9.096	2.294	6.056	-14.986	13.496
79P2N2	14.111	4.366	1.952	2.872	-33.247	12.617
80P3N2	13.417	3.275	2.030	3.162	-31.427	12.416
81P2N2	17.496	7.405	2.212	2.498	-55.285	14.992
84P2N2	19.336	9.093	2.904	1.817	-74.810	15.773
9P1N1	17.496	7.405	2.212	2.498	-55.285	14.992

There was a significant interaction effect between the strains and carbon types as well as strains and carbon sources ($p < 0.0001$). There was also a significant difference between strains ($p < 0.0001$) and between the carbon types ($p < 0.0001$).

There was a significant interaction effect between the soil pH strains were isolated from and carbon sources ($p = 0.0001$). There was also a significant difference between soil pH ($p < 0.0001$) and between the carbon sources ($p < 0.0001$).

Table A 21: ANOVA results for all strains (excluding TA1 and WSM1325), carbon sources and types of carbon, including interaction effects of strains with carbon source and type.

Variate: AUC	Degrees of freedom	Sum squared	Mean squared	F value	p (>F)
Strain	18	1.87×10^{11}	1.04×10^{10}	263.42	< 0.0001
Carbon type	9	1.36×10^{10}	1.51×10^{9}	38.26	< 0.0001
Carbon source	122	6.49×10^{10}	5.32×10^{8}	33.12	< 0.0001
Strain × Carbon type	162	1.24×10^{10}	7.65×10^{7}	1.94	< 0.0001
Strain × Carbon source	2196	1.01×10^{11}	4.58×10^{7}	2.85	< 0.0001

Table A 22: ANOVA results for soil pH (excluding TA1 and WSM1325) and carbon sources, including interaction effects of soil pH with carbon source.

Variate: AUC	Degrees of freedom	Sum squared	Mean squared	F value	p (>F)
Soil pH	1	4.28×10^{10}	4.28×10^{10}	696.79	< 0.0001
Carbon source	122	6.49×10^{10}	5.32×10^{8}	8.654	< 0.0001
Soil pH × Carbon source	122	1.02×10^{10}	8.43×10^{7}	1.358	0.0059

Table A 23: PM01 and PM02 carbon sources which had AUCs lower than the set threshold.

C-Sources (PM01 plate)	C-type
D-Saccharic Acid	Carboxylic acid
D-Alanine	Amino acid
D-Serine	Amino acid
D-Glucuronic Acid	Carboxylic acid
Thymidine	Carbohydrate
D-Aspartic Acid	Amino acid
1,2-Propanediol	Alcohol
a-Keto-Glutaric Acid	Carboxylic acid
a-Keto-Butyric Acid	Carboxylic acid
m-Tartaric Acid	Carboxylic acid
a-Hydroxy Glutaric Acid-G-Lactone	Carboxylic acid
a-Hydroxy Butyric Acid	Carboxylic acid
2-Deoxy Adenosine	Carbohydrate
Citric Acid	Carboxylic acid
D-Threonine	Amino acid
Mucic Acid	Carboxylic acid
Glycolic Acid	Carboxylic acid
Glyoxylic Acid	Carboxylic acid
Tricarballylic Acid	Carboxylic acid
L-Threonine	Amino acid
m-Hydroxy Phenyl Acetic Acid	Carboxylic acid
Tyramine	Amine
L-Galactonic Acid-G-Lactone	Carboxylic acid
Phenylethyl-amine	Amine

C-Sources (PM02 plate)	C-type
Chondroitin Sulfate C	Polymer
a-Cyclodextrin	Polymer
b-Cyclodextrin	Polymer
g-Cyclodextrin	Polymer
Gelatin	Polymer
N-Acetyl-Neuraminic acid	Carboxylic acid
L-Glucose	Carbohydrate
D-Melezitose	Carbohydrate
3-Methylglucose	Carbohydrate
b-Methyl-D-Glucuronic acid	Carboxylic acid
Sedoheptulosan	Carbohydrate
Stachyose	Carbohydrate
N-Acetyl-D-Glucosaminitol	Carbohydrate
d-Amino Valeric acid	Carboxylic acid
Capric acid	Carboxylic acid
Caproic acid	Carboxylic acid
Citraconic acid	Carboxylic acid
D,L-Citramalic acid	Carboxylic acid
2-Hydroxybenzoic acid	Carboxylic acid
4-Hydroxybenzoic acid	Carboxylic acid
g-Hydroxybutyric acid	Carboxylic acid
a-Keto-Valeric acid	Carboxylic acid
Itaconic acid	Carboxylic acid
D-Lactic acid Methyl Ester	Ester
Malonic acid	Carboxylic acid
Oxalic acid	Carboxylic acid
D-Ribono-1,4-Lactone	Carboxylic acid
Sebacic acid	Carboxylic acid
Sorbic acid	Carboxylic acid
D-Tartaric acid	Carboxylic acid
L-Tartaric acid	Carboxylic acid
Acetamide	Amide
N-Acetyl-L-Glutamic acid	Amino acid
Glycine	Amino acid
L-Homoserine	Amino acid
L-Isoleucine	Amino acid
L-Leucine	Amino acid
L-Methionine	Amino acid
L-Phenylalanine	Amino acid
L-Valine	Amino acid
sec-Butylamine	Amine
Putrescine	Amine
2,3-Butanediol	Alcohol
2,3-Butanone	Alcohol
3-Hydroxy-2-butanone	Alcohol

Table A 24: Heat map of AUCs for strains utilising each carbon source (blue = greater AUC, red = smaller AUC). Carbon sources arranged from highest to lowest average AUCs.

Carbon	AUC_Ave	TA1	WSM1325	81P2N2	80P3N2	47P2N1	78P2N2	41P1N2	84P2N2	5P2N2	1P1N1	79P2N2	41P3N2	5P2N1	43P2N2	73P1N2	5P1N2	74P3N2	6P1N2	46P2N1	4P1N1	9P1N1
D-Glucosamine	53092	51763	40637	38284	43395	44026	46332	46690	47351	47453	47494	47907	48623	49966	50458	50574	50896	52625	73448	75658	79274	82073
Laminarin	46092	42104	31196	40195	25382	28234	37049	43687	30853	44639	39950	46407	43524	47583	36906	43101	48314	50419	74520	62764	84116	66985
D-Arabinose	45052	43554	36106	34913	35561	34026	40619	45048	38626	44563	43101	43344	42592	48166	45749	38152	41017	48419	57584	59534	62920	62501
Dihydroxyacetone	45014	47508	40613	39379	39580	42313	41768	43892	43872	50088	43446	45371	45280	46602	43680	43021	45467	46810	47786	47758	50050	51002
Acetoacetic_acid	44934	29427	31652	21590	26565	47597	33314	39674	44944	40581	60139	35038	40751	40414	52259	53571	41901	36973	63974	64872	67580	70789
D-Arabitol	44766	43420	36397	38514	29874	28450	41895	46443	30849	47108	46555	48872	46543	47667	41863	41584	40022	50189	56754	57231	60261	59586
Pectin	43964	46141	35746	44222	30861	29383	39663	49092	31969	50310	36034	50573	50764	51706	39845	42300	39380	52882	42384	57459	41725	60810
L-Lyxose	43858	44438	40168	38138	37059	42064	42083	42901	44149	42027	46738	42014	43315	42183	45584	44783	43340	46115	47504	47335	47520	51567
L-Arabinose	43037	35357	35407	39981	30224	32870	33827	44719	34485	45449	39103	46954	45462	50045	37479	42534	41035	50415	45866	60028	57913	54623
D-Xylose	42994	38301	36425	39523	33076	33495	36327	45873	35641	48048	40289	47846	45524	49299	41730	43470	38686	51191	42473	53741	51359	50555
5-Keto-D-Gluconic_acid	42725	38134	35838	30655	30383	34140	38639	43235	37214	43875	41195	43161	41715	43201	41247	36214	41270	41302	57290	58401	60133	59986
Turanose	42512	43952	34394	38578	26895	26489	37784	45703	30382	44750	41436	47223	45539	47322	40725	42423	30576	50749	50462	50766	56993	59612
Hydroxy-L-Proline	41686	33964	24810	20841	22676	25114	32450	36505	25244	33689	34847	38419	34495	40270	35141	34880	48196	43520	81174	60635	93165	75369
Oxalomalic_acid	41376	44217	39231	35603	31743	39713	40990	38857	39627	39466	45017	40708	39248	35903	47165	44961	41020	39121	42851	49151	41968	52326
Pyruvic_acid	40809	40640	36047	31577	27783	32230	34409	44993	31811	44605	38973	47368	44160	45444	38923	39919	30401	50369	39922	50071	50073	57262
L-Arabitol	40781	39013	34052	22109	27329	28760	38531	43917	31162	39191	44901	44171	41996	43003	42487	36332	39264	43758	48791	52779	55841	59022
Palatinose	40589	37860	34363	33592	28205	28465	36857	41965	30440	44236	38042	42619	42311	40946	39199	36772	34323	45834	51104	51013	57636	56598
D-Fucose	40216	40817	35792	34446	28419	33314	40754	39605	33822	40657	40537	39472	35701	41583	41459	39194	35838	46517	48636	47710	48844	51417
L-Alaninamide	39824	42800	32488	36827	29946	35790	30804	32361	34930	24980	56969	34149	31385	25983	56704	32695	34604	32284	47651	55349	59995	67602
L-Sorbose	39558	39206	35639	21261	28756	32928	38156	45758	32924	45047	40055	47035	44988	41687	43032	35112	30491	33869	46617	47948	48774	51430
b-Methyl-D-Galactoside	39425	38730	29697	23165	24078	26440	36551	41188	30346	39852	43447	42988	41765	44261	43887	33632	35618	46905	47230	47333	53534	57283
a-Methyl-D-Glucoside	39333	40515	35013	25709	29957	35408	38420	40854	34199	39719	42243	43044	37111	32636	45481	34824	32816	43373	45773	48531	49316	51059
b-Methyl-D-Xyloside	39262	42010	36944	24470	27479	28138	38673	41484	31671	41238	41677	43106	40862	43496	39687	34403	29723	46367	48635	47609	53120	43716
Maltitol	39084	39479	34321	34041	25045	29786	39684	37373	30696	40239	37450	37607	34128	32934	44599	41291	28683	42846	50811	52431	54565	52747
D-Lactitol	39070	39714	27435	35055	26032	26332	35551	41348	28039	39354	38879	44291	40676	46150	39962	30615	35414	46292	46017	48301	50744	54276
D-Ribose	39030	36558	33254	35984	31867	33094	33632	39494	35357	39431	37502	40204	39533	39724	37949	43783	35429	38053	40125	43905	51628	53124
Xylitol	38972	37812	30032	18827	23871	24970	33909	43003	28332	39960	38530	45306	40979	41825	40285	45998	35872	44865	48502	47225	53937	54365
Acetic_acid	38914	36491	29214	37832	25839	25078	32462	42703	29139	40684	37619	46085	42844	42587	41248	38298	24025	45939	44366	53676	43879	57177
L-Rhamnose	38819	37433	31747	34546	23152	28978	30137	42166	29527	41041	37831	44033	43186	45832	37047	42523	26336	49469	38417	50243	48931	52629
D-Tagatose	38762	38706	30402	29780	26991	25597	33186	46605	29782	45793	37545	44150	44938	46768	42427	33312	30654	39136	42961	46027	50877	48370
D-Sorbitol	38613	29920	26983	34212	19960	26114	35380	45516	27171	45551	35118	46459	45117	45893	31647	43337	24357	46991	48179	49611	54227	49124
b-D-Allose	38484	39601	34030	22020	25999	31283	36267	45958	30913	41172	40857	45096	43102	40732	42124	36865	31000	32995	43737	46859	45840	51723
L-Histidine	38435	37641	34368	19635	26230	31663	40238	43360	30624	43161	36956	41142	39982	46234	39788	24783	33041	43791	45634	47276	51350	50240
Dextrin	38387	35427	28029	19936	26103	28898	32959	46310	32149	45163	42026	47140	41469	41564	41453	37759	25794	42147	43559	50199	46221	51823
L-Malic_acid	38310	36894	31486	39740	25297	29723	27875	43508	30603	44293	36175	47784	44455	47013	33640	29492	28641	51397	32426	50249	37946	55873
D-Fructose	38289	35880	29232	36330	20453	24636	27966	44611	25118	43433	33318	46633	44125	47771	33451	46872	24909	50300	34638	50869	49211	54319
Adonitol	37890	38581	31460	23162	24674	24491	30204	43380	25735	41366	33566	44466	43371	43147	34835	46694	27358	49352	35481	49862	51557	52947
Glycerol	37802	38575	29825	32628	20161	25136	32241	45665	23488	43904	35936	44767	44054	46615	32798	27009	23541	43623	42014	53582	53662	54628
D,L-Carnitine	37787	42092	35091	19665	23993	29367	38736	39037	30247	36488	38823	38293	37209	38726	43105	34050	25362	43933	47641	48250	51495	51926

g-Amino-N-Butyric_acid	37456	33350	26101	35909	24089	28586	31750	38158	29746	39024	37866	39368	37842	41804	38561	44855	35320	46447	39197	44688	45050	48866
Methyl_Pyruvate	37446	39648	31272	22784	26532	26676	31203	42573	25754	40836	35397	43446	41761	43814	38131	40514	28124	45973	38737	47413	46350	49428
N-Acetyl-D-Galactosamine	37427	33626	28930	38560	24393	28604	33047	38698	29187	36136	36605	41854	38679	42729	35460	31169	34371	46686	41122	47459	46246	52415
a-D-Glucose	37394	31852	28920	36603	21057	25450	26914	43231	27322	43743	31542	46113	42532	46223	30884	46844	26596	50193	32568	50656	49905	46133
Succinic_acid	37349	34558	31103	37054	22989	27038	27869	42570	27097	43279	35951	46320	42774	46360	34721	26391	22401	49334	34947	50177	45848	55549
i-Erythritol	37334	37537	35952	24049	29721	32547	39316	35119	34449	32224	33627	35467	33799	40752	35608	42991	28801	49980	48887	41478	38939	52766
Quinic_acid	37259	38113	29329	35576	24880	24007	37731	41521	26959	41815	37126	42977	39036	29504	41738	27191	28267	45496	44251	43680	50712	52526
D-Mannose	37201	32166	28916	35423	22348	28009	27790	43925	28900	42596	34940	47606	43794	46072	30723	45891	27472	50030	32874	49050	40707	41995
Gentiobiose	37147	33847	28548	38036	23081	26084	32746	36025	28260	35897	34188	44881	36652	45999	31587	38793	30578	48073	40769	43538	48931	53580
D-Mannitol	37037	30805	28717	37355	21938	28293	27606	44360	28622	42992	32129	46117	43018	44945	28971	46053	25795	49916	31998	48641	44045	45456
Arbutin	36997	39973	26761	36563	25291	24610	28039	39440	25927	34855	36171	38216	37142	39078	36685	37523	28747	46685	44634	48298	49679	52625
a-Methyl-D-Mannoside	36990	36043	31252	20298	25363	32757	33039	42996	29999	39219	39863	42003	40090	40934	39999	30191	32107	36391	40863	45619	44997	52768
Amygdalin	36941	33852	27016	29292	21195	27869	30961	39565	29770	36441	38032	43130	39216	43073	36031	32551	34073	42988	44800	44266	48798	52846
Sucrose	36751	31826	28532	37584	21991	27623	25883	43859	28506	43771	32996	46786	43184	47190	30649	42677	27069	50432	31650	49508	38021	42035
Salicin	36735	33893	27173	34712	20540	25685	32186	38199	27852	39183	33234	42023	38848	44071	36101	36233	31969	45660	39644	45657	46055	52521
Inulin	36692	35868	30318	22963	25381	32540	35718	38606	32345	36663	41131	38019	36755	34668	40915	35349	33530	36444	44496	44359	44853	49612
D-Raffinose	36557	31471	27095	24490	21735	26684	33250	39784	28102	35213	36018	46203	38077	45457	32059	30673	31414	46625	44369	46799	51345	50827
b-Hydroxybutyric_acid	36344	36621	28343	26293	24351	30663	30547	40303	31610	38845	35920	38150	34249	34499	41812	38991	29538	37182	42052	45712	47056	50486
D-Trehalose	36270	31160	27423	37874	21751	28521	26287	42583	29105	42896	33407	46413	44461	45134	29797	43212	25967	50359	31438	43312	40311	40248
Glycogen	36149	33220	28997	21130	26252	34109	30724	39935	35088	36797	41507	38932	36924	35134	45600	32281	31749	34977	42989	41530	43495	47762
Tween80	35948	28480	26509	27486	18642	29184	26723	39077	27687	36493	38848	41270	38286	35641	41031	37712	28538	38021	43218	48757	50147	53158
D-Malic_acid	35633	33534	27343	31728	24057	29286	28813	40956	29901	38959	35279	43789	41855	43935	39607	30831	27318	34954	33542	46442	40980	45193
Formic_acid	35614	31467	22216	21241	23680	27599	27220	39805	26501	40593	29930	37594	37898	46303	31161	35196	27065	42819	44850	50826	51525	52408
Melibionic_acid	35562	32772	27755	22837	24260	33035	37419	37641	33214	35719	35801	35705	30655	41569	34932	30179	31304	39069	48230	36892	51313	46508
L-Pyroglutamic_acid	35506	33255	23829	38191	28144	30590	36296	34704	31433	48615	31265	34262	46870	32063	44412	28866	24926	30851	44194	49006	34025	39830
Maltose	35388	29867	26720	36157	20747	26310	24521	42309	26990	42127	31139	46051	42261	46174	30794	43777	25553	48515	30276	45608	36836	40415
D-Galactose	35202	29761	26484	35208	19445	26864	25197	39697	28890	40287	33606	46179	42070	44295	28980	43586	25997	50378	31499	41576	39756	39492
Butyric_acid	35116	32147	28016	28623	24535	29831	37052	34762	29727	36161	35893	37760	33941	33362	37678	37587	29418	37759	43065	45770	39936	44417
Tween40	35099	25849	24581	28292	17864	27287	24750	41081	27933	39687	38127	42338	39946	36342	39163	38766	26885	37831	38796	45142	46266	50156
3-0-b-D-Galactopyranosyl-D-Arabinose	35062	33802	28415	24208	25713	33534	29796	30825	35441	33177	39607	32201	31529	34902	38684	32300	35372	33690	43825	44674	45120	49492
D-Psicose	35022	29426	26467	25645	19260	26810	24811	39569	28648	41003	33512	42418	40946	34195	30621	33928	28323	38388	37851	52150	54450	47050
Fumaric_acid	34932	32756	28806	35495	18990	26050	24778	41287	25556	41784	33877	43699	41250	44253	33358	24632	21318	48568	26634	46429	39862	54200
Mannan	34724	33272	28790	23393	25933	32574	31754	38219	33003	35323	37604	36854	35525	35148	41643	35229	32151	34401	40767	37759	37247	42611
Maltotriose	34684	29256	25452	31031	20547	27816	23901	39784	28726	40023	32789	43734	37926	43388	32072	44196	26532	45879	31113	43374	39206	41611
2-Deoxy-D-Ribose	34670	37196	35264	29089	30376	32383	32239	37112	33571	35626	36494	35498	32826	35897	35183	32260	32537	33570	36458	34842	37066	42588
a-D-Lactose	34649	30842	26020	34514	20295	24143	24924	42239	25930	40563	30675	45572	41733	44797	29928	28867	24875	47324	30952	42783	42932	47729
Lactulose	34553	30263	25383	30322	21006	26408	25850	40878	28128	37746	32750	43508	41669	45604	32074	33979	25976	47638	33230	41948	41156	40091
Glucuronamide	34535	33682	29963	24509	23898	30273	32261	37626	29545	36783	37209	38994	37876	36235	35283	30613	29490	34796	39098	43438	40486	43185
Uridine	34484	32175	26512	26489	21136	27382	26158	36407	29407	36583	33409	41767	37302	40432	30810	40030	25969	46608	35408	43265	43512	43402
Tween20	34431	26923	24245	30804	18813	27832	24711	39822	24541	38250	34519	42656	39465	37384	39566	35608	24781	40866	34666	45867	44116	47610
D-Cellobiose	34378	31043	26213	36122	21326	25601	25137	41572	27456	39080	30831	44148	40047	44373	26906	41234	27252	47533	31844	39904	35289	39034
D,L-Malic_acid	34270	31782	27123	34587	19175	25425	21924	41026	25356	40830	33592	43091	39978	44214	32751	27058	22297	48380	26133	48352	31543	55046

a-Methyl-D-Galactoside	34173	31260	25448	31454	19141	23901	24348	41351	25774	40635	30882	44549	40918	43060	33078	29932	24287	48108	31467	49093	38905	40039
b-Methyl-D-Glucoside	34136	29647	25519	33171	19080	23253	23759	37587	24853	35451	31461	41889	39112	42831	32364	41815	25987	45797	34570	47468	41012	40232
L-Glutamic_acid	34109	30146	24959	36207	20957	27189	27381	35387	27740	36311	32749	40005	38776	41481	33024	31824	25607	46950	33096	46284	39032	41176
L-Glutamine	33990	27637	21495	34648	22780	24913	26329	32527	25816	42489	31890	40705	41865	41907	35484	36590	24366	46440	35665	46840	35581	37824
D-Galacturonic_acid	33990	32384	27559	27291	23472	29573	28623	37194	31876	35985	36384	39583	37777	36891	33800	33425	28107	35676	35266	40914	39549	42454
L-Fucose	33989	32499	26707	34058	18406	24278	24809	35372	25925	32470	33031	39664	38230	42910	33801	42026	24253	46748	32517	42513	42578	40975
m-Inositol	33987	29385	23869	34973	19306	22908	24339	41372	24332	41021	29163	44336	40396	44527	26519	38224	24030	49065	30755	40049	47384	37768
N-Acetyl-D-Glucosamine	33904	27738	21497	32085	18374	25110	25738	40153	26830	36375	31240	45275	41541	44759	29922	42812	22126	48871	29217	42914	38098	41320
D-Melibiose	33841	29337	25442	36466	20574	25579	25486	40644	26821	37367	31662	44350	39608	42991	28569	29708	26208	47463	32257	45609	36121	38401
N-Acetyl-b-D-Mannosamine	33819	31736	25345	24408	20778	24103	28513	37960	25504	35297	31992	38681	43680	33302	32085	30314	26654	47424	40731	45149	45074	41467
Succinamic_acid	33541	33571	24846	22327	22309	30411	28703	40193	29947	37766	33506	39240	36042	35016	39159	30987	26876	37004	33085	38473	38619	46277
L-Lactic_acid	33055	34369	26509	33562	22244	26987	26274	35457	26250	30741	29576	36428	34588	37470	33382	40834	25545	39509	35037	39221	41459	38719
L-Alanine	32897	26986	27643	21609	19533	22358	31555	29103	22113	38927	32084	32629	35539	39609	27099	33166	22634	46378	42620	43920	49110	46222
D-Fructose-6-Phosphate	32848	28029	24449	25689	19452	24712	25548	35949	26525	42447	30954	43094	42075	43905	33383	27275	25601	32669	32087	41532	40137	44307
L-Asparagine	32784	26277	21328	29088	18483	23620	30637	32069	25325	45687	30128	37458	30079	42083	26528	32187	23950	43168	42196	48388	42229	37550
L-Proline	32780	27556	24129	32627	19764	24545	24322	37077	25742	36344	30051	41049	37511	42588	27841	45336	23727	47980	30244	37766	35344	36844
D-Galactonic_acid-G-Lactone	32662	27804	23063	19349	21449	24544	24739	34789	25734	29833	30958	35781	33258	36627	36274	27312	32267	35850	38925	46699	47545	53091
Dulcitol	32615	29037	25788	21135	20332	26937	26245	38660	27399	38683	32489	42694	38536	37580	32361	29673	25674	32227	33401	42950	41604	41510
Neg_Cont_PM2	32494	29492	25389	21499	21672	27392	29579	35075	28629	35087	33550	36626	33725	32414	33982	31317	29990	32770	39375	39875	41118	43823
Mono_Methyl_Succinate	32397	33999	26371	30735	21822	22378	28381	36248	23180	37951	32019	37556	35666	36740	31285	31167	22307	42357	27653	43021	34272	45233
L-Alanyl-Glycine	32275	29172	24321	28999	19562	22312	23893	38531	23326	35867	28549	41665	38778	36758	32255	36623	23444	44785	27412	39402	35680	46437
L-Ornithine	32272	28885	23888	34259	21099	25944	29240	32461	27208	32037	31305	35038	31227	40796	31426	33824	26165	41752	36754	37361	36143	40894
D-Gluconic_acid	32049	28699	25367	19441	17855	24865	23428	40955	26619	37913	31857	41619	39916	36723	30306	29198	23960	39160	31312	39914	41617	42316
Inosine	31747	29226	24730	29879	21926	25370	24927	31892	26735	32494	29493	33179	31021	32060	28163	42702	26781	43235	32585	38167	41374	40757
D-Glucose-1-Phosphate	31454	25933	22360	31836	17171	23252	22171	35408	24794	42268	28876	47618	43770	48183	28959	28791	21726	30591	28279	35545	34282	38728
2-Aminoethanol	31446	31394	26612	25511	23704	27780	27753	32097	28993	31248	31340	34967	33349	32515	30756	31180	26766	34177	32611	41798	37900	37910
Adenosine	31406	29185	23003	23252	21058	26568	23540	29125	28262	29236	31529	27039	28787	37178	30176	41860	27937	41880	34841	41842	42273	40953
Glycyl-L-Glutamic_acid	31309	28233	24852	24514	21442	25203	27176	33383	27288	32321	32082	36884	33783	31211	29610	33011	26413	41749	32527	36364	39342	40099
Bromo_Succinic_acid	31213	27326	25099	19680	19353	23221	21219	38636	23190	39975	27487	43360	38684	42022	32050	21647	22293	46633	24684	41530	29964	47429
D-Glucose-6-Phosphate	31174	24825	21796	33377	17706	22716	23541	30578	24239	40573	28745	49491	44593	49467	26361	26171	22013	29207	28475	37062	35900	37819
D-Glucosaminic_acid	31031	27593	22729	16242	17008	24466	21761	41453	25514	36511	31884	45716	40707	45956	28203	27158	21953	34294	28572	38414	34853	40665
L-Lysine	30950	29615	24554	32667	22168	26680	29831	27103	27621	28389	30850	27052	25969	25575	32896	38017	26694	43550	36832	36735	36149	41001
Glycyl-L-Aspartic_acid	30884	27474	22995	22804	19371	24869	25469	35303	26880	33704	31288	39635	35398	34185	29004	32258	23228	39700	30373	35464	39005	40164
L-Arginine	30816	26840	22839	30194	17813	23508	26136	31317	25805	29727	32181	35911	30632	36977	32400	32733	22783	40705	32946	34748	37329	43602
L-Serine	30801	29063	25137	19987	20694	27457	25089	32898	28926	29837	33436	36116	32139	39455	30257	30462	25488	41580	30406	34290	33909	40199
D,L-Octopamine	30765	26921	23244	21138	21123	25281	28858	29995	26395	36888	34601	46261	27904	37643	30549	26428	23850	27914	32694	35141	50225	33007
Glycyl-L-Proline	30530	26836	23727	22191	20186	25136	25231	35361	26953	34776	30986	36947	34153	34742	29511	32045	23278	39295	30370	33739	36699	38979
p-Hydroxy_Phenyl_Acetic_acid	30506	24495	22073	20708	20636	24574	29847	27840	24892	39621	31727	41798	26066	40444	25353	26692	25949	30056	37121	43651	45581	31499
L-Aspartic_acid	30441	30285	23468	36549	19020	22938	26193	25034	23415	42478	28363	27013	29348	25129	27460	28436	22876	48024	34019	43768	34973	40482
Mucic_acid	29880	28630	22857	20622	18591	23514	22507	34217	25957	36212	28931	41467	36750	36760	28762	27095	23766	33246	28502	37577	34831	36686
D,L-a-Glycerol-Phosphate	29846	26954	22274	19482	17738	23413	22265	35397	25058	28996	29929	38115	35511	42806	28478	29221	22635	37496	29077	37490	34455	39967
Neg_Cont_PM1	29710	27097	22762	20158	19856	23995	26671	33858	25709	32802	31574	35569	32904	30872	28972	28861	23734	30606	33666	35726	38497	40023

References

Abd-Alla, M.H., El-Enany, A.-W.E., Nafady, N.A., Khalaf, D.M. and Morsy, F.M. (2014) Synergistic interaction of *Rhizobium leguminosarum* bv. *viciae* and arbuscular mycorrhizal fungi as a plant growth promoting biofertilizers for faba bean (*Vicia faba* L.) in alkaline soil. *Microbiological Research* **169**, 49-58.

Ahmad, F., Ahmad, I. and Khan, M. (2008) Screening of free-living rhizospheric bacteria for their multiple plant growth promoting activities. *Microbiological Research* **163**, 173-181.

Alexandre, A. and Oliveira, S. (2013) Response to temperature stress in rhizobia. *Critical Reviews in Microbiology* **39**, 219-228.

Amarger, N. (2001) Rhizobia in the field. *Advances in Agronomy* **73**, 109-168.

Anderson, M., Gorley, R.N. and Clarke, R.K. (2008) *PERMANOVA+ for Primer: Guide to Software and Statisticl Methods*: Primer-E Limited.

Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L., Sanders, N.J., Cornell, H.V., Comita, L.S. and Davies, K.F. (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology Letters* **14**, 19-28.

Andersson, S., Nilsson, S.I. and Saetre, P. (2000) Leaching of dissolved organic carbon (DOC) and dissolved organic nitrogen (DON) in mor humus as affected by temperature and pH. *Soil Biology and Biochemistry* **32**, 1-10.

Andrade, D., Murphy, P. and Giller, K. (2002) The diversity of *Phaseolus*-nodulating rhizobial populations is altered by liming of acid soils planted with *Phaseolus vulgaris* L. in Brazil. *Applied and Environmental Microbiology* **68**, 4025-4034.

Andrew, C. (1978) Legumes and Acid Soils. In *Limitations and Potentials for Biological Nitrogen Fixation in the Tropics* ed. Döbereiner J., B.R.H., Hollaender A., Franco A.A., Neyra C.A., Scott D.B. pp.135-160: Springer, Boston, MA.

Andrews, M., De, S.M., James, E.K., Stępkowski, T., Hodge, S., Simon, M.F. and Young, J.P.W. (2018) Horizontal Transfer of Symbiosis Genes within and Between Rhizobial Genera: Occurrence and Importance. *Genes* **9**.

Andrews, M., James, E.K., Sprent, J.I., Boddey, R.M., Gross, E. and dos Reis Jr, F.B. (2011) Nitrogen fixation in legumes and actinorhizal plants in natural ecosystems: values obtained using ^{15}N natural abundance. *Plant Ecology & Diversity* **4**, 131-140.

Bååth, E. and Anderson, T.-H. (2003) Comparison of soil fungal/bacterial ratios in a pH gradient using physiological and PLFA-based techniques. *Soil Biology and Biochemistry* **35**, 955-963.

Baev, N., Endre, G., Petrovics, G., Banfalvi, Z. and Kondorosi, A. (1991) Six nodulation genes of *nod* box locus 4 in *Rhizobium meliloti* are involved in nodulation signal production: *nodM* codes for D-glucosamine synthetase. *Molecular and General Genetics* **228**, 113-124.

Ballard, R., Charman, N., McInnes, A. and Davidson, J. (2004) Size, symbiotic effectiveness and genetic diversity of field pea rhizobia (*Rhizobium leguminosarum* bv. *viciae*) populations in South Australian soils. *Soil Biology and Biochemistry* **36**, 1347-1355.

Ballard, R., Craig, A. and Charman, N. (2002) Nodulation and growth of pasture legumes with naturalised soil rhizobia. 2. Balansa clover (*Trifolium michelianum* Savi). *Australian Journal of Experimental Agriculture* **42**, 939-944.

Ballard, R., Shepherd, B. and Charman, N. (2003) Nodulation and growth of pasture legumes with naturalised soil rhizobia. 3. Lucerne (*Medicago sativa* L.). *Australian Journal of Experimental Agriculture* **43**, 135-140.

Bano, A. and Fatima, M. (2009) Salt tolerance in *Zea mays* (L). following inoculation with *Rhizobium* and *Pseudomonas*. *Biology and Fertility of Soils* **45**, 405-413.

Barea, J., Navarro, E. and Montoya, E. (1976) Production of plant growth regulators by rhizosphere phosphate-solubilizing bacteria. *Journal of Applied Microbiology* **40**, 129-134.

Barnett, M.J. and Fisher, R.F. (2006) Global gene expression in the rhizobial-legume symbiosis. *Symbiosis* **42**, 1-24.

Barny, M.A., Schoonejans, E., Economou, A., Johnston, A.W. and Downie, J.A. (1996) The C-terminal domain of the *Rhizobium leguminosarum* chitin synthase NodC is important for function and determines the orientation of the N-terminal region in the inner membrane. *Molecular Microbiology* **19**, 443-453.

Barret, M., Morrissey, J.P. and O'Gara, F. (2011) Functional genomics analysis of plant growth-promoting rhizobacterial traits involved in rhizosphere competence. *Biology and Fertility of Soils* **47**, 729.

Bashan, Y. and De-Bashan, L. (2005) Plant growth-promoting. In *Encyclopedia of soils in the environment* ed. Hillel, D. pp.103-115. Oxford, U.K.: Elsevier.

Bathe, S., Achouak, W., Hartmann, A., Heulin, T., Schloter, M. and Lebuhn, M. (2006) Genetic and phenotypic microdiversity of *Ochrobactrum* spp. *FEMS Microbiology Ecology* **56**, 272-280.

Berg, G. and Smalla, K. (2009) Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiology Ecology* **68**, 1-13.

Bertin, C., Yang, X. and Weston, L.A. (2003) The role of root exudates and allelochemicals in the rhizosphere. *Plant and Soil* **256**, 67-83.

Black, M., Moolhuijzen, P., Chapman, B., Barrero, R., Howieson, J., Hungria, M. and Bellgard, M. (2012) The genetics of symbiotic nitrogen fixation: Comparative genomics of 14 rhizobia strains by resolution of protein clusters. *Genes* **3**, 138-166.

Blagodatskaya, E.V. and Anderson, T.-H. (1998) Interactive effects of pH and substrate quality on the fungal-to-bacterial ratio and qCO2 of microbial communities in forest soils. *Soil Biology and Biochemistry* **30**, 1269-1274.

Bochner, B.R. (2008) Global phenotypic characterization of bacteria. *FEMS Microbiology Reviews* **33**, 191-205.

Bochner, B.R., Gadzinski, P. and Panomitros, E. (2001) Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Research* **11**, 1246-1255.

Booth, I.R. (1985) Regulation of cytoplasmic pH in bacteria. *Microbiological Reviews* **49**, 359.

Borglin, S., Joyner, D., DeAngelis, K.M., Khudyakov, J., D'haeseleer, P., Joachimiak, M.P. and Hazen, T. (2012) Application of phenotypic microarrays to environmental microbiology. *Current Opinion in Biotechnology* **23**, 41-48.

Brockwell, J. (1963) Accuracy of a plant-infection technique for counting populations of *Rhizobium trifolii*. *Applied Microbiology* **11**, 377-383.

Brockwell, J., Pilka, A. and Holliday, R. (1991) Soil pH is a major determinant of the numbers of naturally occurring *Rhizobium meliloti* in non-cultivated soils in central New South Wales. *Animal Production Science* **31**, 211-219.

Bruning, B. and Rozema, J. (2013) Symbiotic nitrogen fixation in legumes: Perspectives for saline agriculture. *Environmental and Experimental Botany* **92**, 134-143.

Bulgarelli, D., Garrido-Oter, R., Münch, P.C., Weiman, A., Dröge, J., Pan, Y., McHardy, A.C. and Schulze-Lefert, P. (2015) Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host & Microbe* **17**, 392-403.

Campbell, C., Grayston, S. and Hirst, D. (1997) Use of rhizosphere carbon sources in sole carbon source tests to discriminate soil microbial communities. *Journal of Microbiological Methods* **30**, 33-41.

Caradus, J., Woodfield, D. and Stewart, A. (1996) Overview and vision for white clover. *Special Publication - Agronomy Society of New Zealand*, 1-6.

Caradus, J.B., JL; Hay, M (1989) Fifty Years of White Clover Research in New Zealand. In *Proceedings of the New Zealand Grassland Association*. pp.25-39.

Chapman, E.E.V., Dave, G. and Murimboh, J.D. (2013) A review of metal (Pb and Zn) sensitive and pH tolerant bioassay organisms for risk screening of metal-contaminated acidic soils. *Environmental Pollution* **179**, 326-342.

Charlton, D. (2012) Pastures - Creating pastures: Te Ara - the Encyclopedia of New Zealand.

Charman, N., Ballard, R.A., Humphries, A.W. and Auricht, G.C. (2008) Improving lucerne nodulation at low pH: contribution of rhizobial and plant genotype to the nodulation of lucerne seedlings growing in solution culture at pH 5. *Australian Journal of Experimental Agriculture* **48**, 512-517.

Cheng, W., Zhang, Q., Coleman, D.C., Carroll, C.R. and Hoffman, C.A. (1996) Is available carbon limiting microbial respiration in the rhizosphere? *Soil Biology and Biochemistry* **28**, 1283-1288.

Clarke, K. and Ainsworth, M. (1993) A method of linking multivariate community structure to environmental variables. *Marine Ecology - Progress Series* **92**, 205-205.

Clarke, K. and Gorley, R. (2015) PRIMER v7: User manual/tutorial. *Plymouth, UK: PRIMER-E Ltd*.

Clarke, K.R. (1993) Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* **18**, 117-143.

Connell, J.H. (1978) Diversity in tropical rain forests and coral reefs. *Science* **199**, 1302-1310.

Cooper, J. (2007) Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue. *Journal of Applied Microbiology* **103**, 1355-1365.

De Freitas, J., Banerjee, M. and Germida, J. (1997) Phosphate-solubilizing rhizobacteria enhance the growth and yield but not phosphorus uptake of canola (*Brassica napus* L.). *Biology and Fertility of Soils* **24**, 358-364.

de Zelicourt, A., Al-Yousif, M. and Hirt, H. (2013) Rhizosphere microbes as essential partners for plant stress tolerance. *Molecular Plant* **6**, 242-245.

Deaker, R., Roughley, R.J. and Kennedy, I.R. (2004) Legume seed inoculation technology—a review. *Soil Biology and Biochemistry* **36**, 1275-1288.

Delestre, C., Laugraud, A., Ridgway, H., Ronson, C., O'Callaghan, M., Barrett, B., Ballard, R., Griffiths, A., Young, S. and Blond, C. (2015) Genome sequence of the clover symbiont *Rhizobium leguminosarum* bv. *trifolii* strain CC275e. *Standards in Genomic Sciences* **10**, 121.

Dénarié, J., Debelle, F. and Rosenberg, C. (1992) Signaling and host range variation in nodulation. *Annual Reviews in Microbiology* **46**, 497-531.

Dennis, P.G., Miller, A.J. and Hirsch, P.R. (2010) Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiology Ecology* **72**, 313-327.

Denton, M., Coventry, D., Murphy, P., Howieson, J. and Bellotti, W. (2002) Competition between inoculant and naturalised *Rhizobium leguminosarum* bv. *trifolii* for nodulation of annual clovers in alkaline soils. *Crop and Pasture Science* **53**, 1019-1026.

Dilworth, M., Howieson, J., Reeve, W., Tiwari, R. and Glenn, A. (2001) Acid tolerance in legume root nodule bacteria and selecting for it. *Animal Production Science* **41**, 435-446.

Divito, G.A. and Sadras, V.O. (2014) How do phosphorus, potassium and sulphur affect plant growth and biological nitrogen fixation in crop and pasture legumes? A meta-analysis. *Field Crops Research* **156**, 161-171.

Dodt, M., Roehr, J.T., Ahmed, R. and Dieterich, C. (2012) FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. *Biology* **1**, 895-905.

Dorfmueller, H.C., Ferenbach, A.T., Borodkin, V.S. and van Aalten, D.M. (2014) A structural and biochemical model of processive chitin synthesis. *Journal of Biological Chemistry*, jbc. M114. 563353.

Earl, C., Ronson, C. and Ausubel, F. (1987) Genetic and structural analysis of the *Rhizobium meliloti fixA*, *fixB*, *fixC*, and *fixX* genes. *Journal of Bacteriology* **169**, 1127-1136.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.

Fan, K., Weisenhorn, P., Gilbert, J.A., Shi, Y., Bai, Y. and Chu, H. (2018) Soil pH correlates with the co-occurrence and assemblage process of diazotrophic communities in rhizosphere and bulk soils of wheat fields. *Soil Biology and Biochemistry* **121**, 185-192.

Ferguson, B.J., Indrasumunar, A., Hayashi, S., Lin, M.H., Lin, Y.H., Reid, D.E. and Gresshoff, P.M. (2010) Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology* **52**, 61-76.

Fierer, N. and Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences* **103**, 626-631.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H. and Caporaso, J.G. (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences* **109**, 21390-21395.

Fischer, H.-M. (1994) Genetic regulation of nitrogen fixation in rhizobia. *Microbiological Reviews* **58**, 352.

Foy, C.D. (1984) Physiological Effects of Hydrogen, Aluminum, and Manganese Toxicities in Acid Soil. *Soil Acidity and Liming*, 57-97.

Fry, J., Wood, M. and Poole, P.S. (2001) Investigation of *myo*-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness. *Molecular Plant-Microbe Interactions* **14**, 1016-1025.

Galtier, N., Gouy, M. and Gautier, C. (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Bioinformatics* **12**, 543-548.

Garbeva, P., Van Veen, J. and Van Elsas, J. (2004) Microbial diversity in soil: selection of microbial populations by plant and soil type and implications for disease suppressiveness. *Annual Review of Phytopathology* **42**, 243-270.

Gascuel, O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular biology and evolution* **14**, 685-695.

Gemell, L., Hartley, E. and Herridge, D. (2005) Point-of-sale evaluation of preinoculated and custom-inoculated pasture legume seed. *Animal Production Science* **45**, 161-169.

Giller, K., Herridge, D. and Sprent, J. (2016) The legume-rhizobia symbiosis and assessing the need to inoculate. In *Working with Rhizobia* eds. Howieson, J.G. and Dilworth, M.J. pp.15-24. Canberra, Australia: Australian Centre for International Agricultural Research (ACIAR).

Gomaa, N.H., Hassan, M.O., Fahmy, G.M., González, L., Hammouda, O. and Atteya, A.M. (2015) Flavonoid profiling and nodulation of some legumes in response to the allelopathic stress of *Sonchus oleraceus* L. *Acta Botanica Brasilica* **29**, 553-560.

Graham, P. (2008) Ecology of the root-nodule bacteria of legumes. In *Nitrogen-fixing Leguminous Symbioses* ed. Dilworth MJ, J.E., Sprent JI, Newton WE. pp.23-58: Springer.

Graham, P.H. and Vance, C.P. (2003) Legumes: importance and constraints to greater use. *Plant Physiology* **131**, 872-877.

Greetham, D. (2014) Phenotype microarray technology and its application in industrial biotechnology. *Biotechnology Letters* **36**, 1153-1160.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**, 307-321.

Haag, A.F., Arnold, M.F., Myka, K.K., Kerscher, B., Dall'Angelo, S., Zanda, M., Mergaert, P. and Ferguson, G.P. (2013) Molecular insights into bacteroid development during Rhizobium–legume symbiosis. *FEMS Microbiology Reviews* **37**, 364-383.

Hardoim, P.R., van Overbeek, L.S. and van Elsas, J.D. (2008) Properties of bacterial endophytes and their proposed role in plant growth. *Trends in microbiology* **16**, 463-471.

Hartman, K., van der Heijden, M.G., Roussely-Provent, V., Walser, J.-C. and Schlaeppi, K. (2017) Deciphering composition and function of the root microbiome of a legume plant. *Microbiome* **5**, 2.

Hastings, A. and Drake, A. (1960) Inoculation and pelleting of clover seed. *New Zealand Journal of Agricultural Research* **101**, 619-621.

He, Z., Deng, Y., Van Nostrand, J.D., Tu, Q., Xu, M., Hemme, C.L., Li, X., Wu, L., Gentry, T.J. and Yin, Y. (2010) GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *The ISME Journal* **4**, 1167-1179.

Hermans, S.M., Buckley, H.L., Case, B.S., Curran-Cournane, F., Taylor, M. and Lear, G. (2017) Bacteria as emerging indicators of soil condition. *Applied and Environmental Microbiology* **83**, e02826-02816.

Herridge, D., Roughley, R. and Brockwell, J. (1984) Effect of rhizobia and soil nitrate on the establishment and functioning of the soybean symbiosis in the field. *Crop and Pasture Science* **35**, 149-161.

Hewitt, A. (1998) New Zealand soil classification. Manaaki Whenua Press, Lincoln, New Zealand.

Hirsch, A.M. (1999) Role of lectins (and rhizobial exopolysaccharides) in legume nodulation. *Current Opinion in Plant Biology* **2**, 320-326.

Howieson, J., Yates, R., O'Hara, G., Ryder, M. and Real, D. (2005) The interactions of *Rhizobium leguminosarum* biovar *trifolii* in nodulation of annual and perennial *Trifolium* spp. from diverse centres of origin. *Animal Production Science* **45**, 199-207.

Howieson, J.G. and Ballard, R.A. (2004) Optimising the legume symbiosis in stressful and competitive environments within southern Australia - some contemporary thoughts. *Soil Biology and Biochemistry* **36**, 1261-1273.

Howieson, J.G. and Ewing, M.A. (1984) Soil acidity and legume nodulation. *Journal of the Department of Agriculture* **25**.

Howieson, J.G., Ewing, M.A. and D'Antuono, M.F. (1988) Selection for acid tolerance in *Rhizobium meliloti*. *Plant and Soil* **105**, 179-188.

Humphries, A., Ballard, R., Charman, N., Venkatanagappa, S., Denton, M., Marshall, E., Hayes, R. and Auricht, G. (2009) Developing lucerne and rhizobia with improved tolerance to acidic soils: A novel approach. In *Proceedings of the 14th Australasian Plant Breeding Conference and 11th Society for the Advancement of Plant Breeding in Asia and Oceania (SABRAO) Congress' Cairns, Australia*.

Hurley, M.A. and Roscoe, M. (1983) Automated statistical analysis of microbial enumeration by dilution series. *Journal of Applied Microbiology* **55**, 159-164.

İnceoğlu, Ö., Salles, J.F. and van Elsas, J.D. (2012) Soil and cultivar type shape the bacterial community in the potato rhizosphere. *Microbial Ecology* **63**, 460-470.

Janczarek, M., Rachwał, K., Marzec, A., Grządziel, J. and Palusińska-Szysz, M. (2015) Signal molecules and cell-surface components involved in early stages of the legume–rhizobium interactions. *Applied Soil Ecology* **85**, 94-113.

Johnson, M., Lee, K. and Scow, K. (2003) DNA fingerprinting reveals links among agricultural crops, soil properties, and the composition of soil microbial communities. *Geoderma* **114**, 279-303.

Johri, J.K., Surange, S. and Nautiyal, C.S. (1999) Occurrence of salt, pH, and temperature-tolerant, phosphate-solubilizing bacteria in alkaline soils. *Current Microbiology* **39**, 89-93.

Jost, L. (2006) Entropy and diversity. *Oikos* **113**, 363-375.

Kaiser, K., Wemheuer, B., Korolkow, V., Wemheuer, F., Nacke, H., Schöning, I., Schrumpf, M. and Daniel, R. (2016) Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Scientific Reports* **6**, 33696.

Kaschuk, G., Kuyper, T.W., Leffelaar, P.A., Hungria, M. and Giller, K.E. (2009) Are the rates of photosynthesis stimulated by the carbon sink strength of rhizobial and arbuscular mycorrhizal symbioses? *Soil Biology and Biochemistry* **41**, 1233-1244.

Keddy, C.J., Greene, J.C. and Bonnell, M.A. (1995) Review of whole-organism bioassays: soil, freshwater sediment, and freshwater assessment in Canada. *Ecotoxicology and Environmental Safety* **30**, 221-251.

Kemmitt, S.J., Wright, D., Goulding, K.W. and Jones, D.L. (2006) pH regulation of carbon and nitrogen dynamics in two agricultural soils. *Soil Biology and Biochemistry* **38**, 898-911.

Kiers, E.T., Rousseau, R.A. and Denison, R.F. (2006) Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. *Evolutionary Ecology Research* **8**, 1077-1086.

Krulwich, T.A., Sachs, G. and Padan, E. (2011) Molecular aspects of bacterial pH sensing and homeostasis. *Nature Reviews Microbiology* **9**, 330.

Kumar, R., Pandey, S. and Pandey, A. (2006) Plant roots and carbon sequestration. *Current Science* **91**, 885-890.

Laguerre, G., Nour, S.M., Macheret, V., Sanjuan, J., Drouin, P. and Amarger, N. (2001) Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts. *Microbiology* **147**, 981-993.

Lammel, D.R., Barth, G., Ovaskainen, O., Cruz, L.M., Zanatta, J.A., Ryo, M., de Souza, E.M. and Pedrosa, F.O. (2018) Direct and indirect effects of a pH gradient bring insights into the mechanisms driving prokaryotic community structures. *Microbiome* **6**, 106.

Lauber, C.L., Hamady, M., Knight, R. and Fierer, N. (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology* **75**, 5111-5120.

Lauber, C.L., Strickland, M.S., Bradford, M.A. and Fierer, N. (2008) The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biology and Biochemistry* **40**, 2407-2415.

Ledgard, S.F. (2001) Nitrogen cycling in low input legume-based agriculture, with emphasis on legume/grass pastures. *Plant and Soil* **228**, 43-59.

Leite, J., Fischer, D., Rouws, L.F.M., Fernandes-Júnior, P.I., Hofmann, A., Kublik, S., Schloter, M., Xavier, G.R. and Radl, V. (2017) Cowpea Nodules Harbor Non-rhizobial Bacterial Communities that Are Shaped by Soil Type Rather than Plant Genotype. *Frontiers in Plant Science* **7**.

Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J.C. and Dénarié, J. (1990) Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* **344**, 781.

Lin, M.-H., Gresshoff, P.M.M. and Ferguson, B.J. (2012) Systemic regulation of soybean nodulation by acidic growth conditions. *Plant Physiology*.

Liu, X.Y., Wang, E.T., Li, Y. and Chen, W.X. (2007) Diverse bacteria isolated from root nodules of *Trifolium*, *Crotalaria* and *Mimosa* grown in the subtropical regions of China. *Archives of Microbiology* **188**, 1-14.

Liu, Y., Wu, L., Baddeley, J.A. and Watson, C.A. (2010) Models of biological nitrogen fixation of legumes. A review. *Agronomy for Sustainable Development*.

Lodwig, E. and Poole, P. (2003) Metabolism of *Rhizobium* bacteroids. *Critical Reviews in Plant Sciences* **22**, 37-78.

Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F.M., Ferrera, I., Sarmento, H., Hingamp, P., Ogata, H., Vargas, C. and Lima-Mendez, G. (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental Microbiology* **16**, 2659-2671.

Long, S.R. (2001) Genes and signals in the *Rhizobium*-legume symbiosis. *Plant Physiology* **125**, 69-72.

López-García, S.L., Vázquez, T.E., Favelukes, G. and Lodeiro, A.R. (2002) Rhizobial position as a main determinant in the problem of competition for nodulation in soybean. *Environmental Microbiology* **4**, 216-224.

Lowther, W. and Kerr, G. (2011) White clover seed inoculation and coating in New Zealand. *Proceedings of the New Zealand Grassland Association* **73**, 93-102.

Lundberg, D.S., Lebeis, S.L., Paredes, S.H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V. and Del Rio, T.G. (2012) Defining the core *Arabidopsis thaliana* root microbiome. *Nature* **488**, 86.

Luo, C., Rodriguez-R, L.M., Johnston, E.R., Wu, L., Cheng, L., Xue, K., Tu, Q., Deng, Y., He, Z. and Shi, J.Z. (2014) Soil microbial community responses to a decade of warming as revealed by comparative metagenomics. *Applied and Environmental Microbiology* **80**, 1777-1786.

Lynch, J. and Whipps, J. (1990) Substrate flow in the rhizosphere. *Plant and Soil* **129**, 1-10.

MacKinnon, P.A., Robertson, J., Scott, D. and Hale, C. (1977) Legume inoculant usage in New Zealand. *New Zealand Journal of Experimental Agriculture* **5**, 35-39.

Marilley, L. and Aragno, M. (1999) Phylogenetic diversity of bacterial communities differing in degree of proximity of *Lolium perenne* and *Trifolium repens* roots. *Applied Soil Ecology* **13**, 127-136.

Marshall, D.J., Bateman, J.D. and Brockwell, J. (1993) Validation of a serial-dilution, plant-infection test for enumerating *Rhizobium leguminosarum* bv. *viciae* and its application for counting rhizobia in acid soils. *Soil Biology and Biochemistry* **25**, 261-268.

Martínez-Hidalgo, P. and Hirsch, A.M. (2017) The Nodule Microbiome: N_2-Fixing Rhizobia Do Not Live Alone. *Phytobiomes* **1**, 70-82.

Maxwell, T., Moir, J. and Edwards, G. (2012) Sulphur and lime response of four adventive annual clovers grown in a New Zealand high country soil under glasshouse conditions. *New Zealand Journal of Agricultural Research* **55**, 47-62.

Mazur, A., Stasiak, G., Wielbo, J., Koper, P., Kubik-Komar, A. and Skorupska, A. (2013) Phenotype profiling of *Rhizobium leguminosarum* bv. *trifolii* clover nodule isolates reveal their both versatile and specialized metabolic capabilities. *Archives of Microbiology* **195**, 255-267.

McArdle, B. and Anderson, M. (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82**, 290-297.

McMurdie, P.J. and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* **8**, e61217.

Mehta, S. and Nautiyal, C.S. (2001) An efficient method for qualitative screening of phosphate-solubilizing bacteria. *Current Microbiology* **43**, 51-56.

Melino, V., Drew, E., Ballard, R., Reeve, W., Thomson, G., White, R. and O'hara, G. (2012) Identifying abnormalities in symbiotic development between *Trifolium* spp. and *Rhizobium leguminosarum* bv. *trifolii* leading to sub-optimal and ineffective nodule phenotypes. *Annals of Botany* **110**, 1559-1572.

Moir, J., Jordan, P., Moot, D. and Lucas, R. (2016) Phosphorus response and optimum pH ranges of twelve pasture legumes grown in an acid upland New Zealand soil under glasshouse conditions. *Journal of Soil Science and Plant Nutrition* **16**.

Moot, D. (2013) An overview of dryland legume research in New Zealand. *Crop and Pasture Science* **63**, 726-733.

Morton, J. and Roberts, A. (1999) Fertiliser use on New Zealand sheep and beef farms. *New Zealand Fertiliser Manufacturer's Association Revised Edition*.

Morton, J.T., Toran, L., Edlund, A., Metcalf, J.L., Lauber, C. and Knight, R. (2017) Uncovering the horseshoe effect in microbial analyses. *MSystems* **2**, e00166-00116.

Munns, D. (1968) Nodulation of *Medicago sativa* in solution culture. *Plant and Soil* **28**, 129-146.

Murty, D., Kirschbaum, M.U., Mcmurtrie, R.E. and Mcgilvray, H. (2002) Does conversion of forest to agricultural land change soil carbon and nitrogen? A review of the literature. *Global Change Biology* **8**, 105-123.

Nadeem, S.M., Ahmad, M., Zahir, Z.A., Javaid, A. and Ashraf, M. (2014) The role of mycorrhizae and plant growth promoting rhizobacteria (PGPR) in improving crop productivity under stressful environments. *Biotechnology Advances* **32**, 429-448.

Nangul, A., Moot, D., Brown, D. and Ridgway, H. (2013) Nodule occupancy by *Rhizobium leguminosarum* strain WSM1325 following inoculation of four annual *Trifolium* species in Canterbury, New Zealand. *New Zealand Journal of Agricultural Research* **56**, 215-223.

Nautiyal, C.S. (1999) An efficient microbiological growth medium for screening phosphate solubilizing microorganisms. *FEMS Microbiology Letters* **170**, 265-270.

Newsome, P. and Pairman, C. (2012) LCDB v3.0 - Land Cover Database. Landcare Research, New Zealand.

Novák, K., Biedermannová, E. and Vondrys, J. (2009) Symbiotic and growth performance of supernodulating forage pea lines. *Crop Science* **49**, 1227-1234.

O'Hara, G. (2001) Nutritional constraints on root nodule bacteria affecting symbiotic nitrogen fixation: a review. *Animal Production Science* **41**, 417-433.

O'Hara, G., Zilli, J., Poole, P. and Hungria, M. (2016) Taxonomy and physiology of rhizobia ed. Dilworth, J.G.H.a.M.J.: Australian Centre for International Agricultural Research.

Oldroyd, G.E., Murray, J.D., Poole, P.S. and Downie, J.A. (2011) The rules of engagement in the legume-rhizobial symbiosis. *Annual review of genetics* **45**, 119-144.

Olivares, F.L., Baldani, V.L., Reis, V.M., Baldani, J.I. and Döbereiner, J. (1996) Occurrence of the endophytic diazotrophs *Herbaspirillum* spp. in roots, stems, and leaves, predominantly of *Gramineae*. *Biology and Fertility of Soils* **21**, 197-200.

Oresnik, I.J., Pacarynuk, L.A., O'Brien, S.A., Yost, C.K. and Hynes, M.F. (1998) Plasmid-encoded catabolic genes in *Rhizobium leguminosarum* bv. *trifolii*: evidence for a plant-inducible rhamnose locus involved in competition for nodulation. *Molecular Plant-Microbe Interactions* **11**, 1175-1185.

Padan, E., Bibi, E., Ito, M. and Krulwich, T.A. (2005) Alkaline pH homeostasis in bacteria: New insights. *Biochimica et Biophysica Acta (BBA - Biomembranes* **1717**, 67-88.

Parfitt, R., Schipper, L., Baisden, W. and Elliott, A. (2006) Nitrogen inputs and outputs for New Zealand in 2001 at national and regional scales. *Biogeochemistry* **80**, 71-88.

Perret, X., Staehelin, C. and Broughton, W.J. (2000) Molecular basis of symbiotic promiscuity. *Microbiology and Molecular Biology Reviews* **64**, 180-201.

Pietri, J.A. and Brookes, P. (2008) Relationships between soil pH and microbial properties in a UK arable soil. *Soil Biology and Biochemistry* **40**, 1856-1861.

Poole, P., Ramachandran, V. and Terpolilli, J. (2018) Rhizobia: from saprophytes to endosymbionts. *Nature Reviews Microbiology*.

Prell, J. and Poole, P. (2006) Metabolic changes of rhizobia in legume nodules. *Trends in Microbiology* **14**, 161-168.

Putnoky, P., Kereszt, A., Nakamura, T., Endre, G., Grosskopf, E., Kiss, P. and Kondorosi, A. (1998) The *pha* gene cluster of *Rhizobium meliloti* involved in pH adaptation and symbiosis encodes a novel type of K^+ efflux system. *Molecular Microbiology* **28**, 1091-1101.

Rai, R., Dash, P.K., Mohapatra, T. and Singh, A. (2012) Phenotypic and molecular characterization of indigenous rhizobia nodulating chickpea in India. *Indian Journal of Experimental Biology* **50**, 340-350.

Ramana, J. (2018) Investigating variation in associations with belowground micro-organisms of historic and current white clover germplasm: Lincoln University.

Rao, D. (2014) Recent Advances in Biological Nitrogen Fixation in Agricultural Systems. In *Proceedings of the Indian National Science Academy*.

Redford, A.J., Bowers, R.M., Knight, R., Linhart, Y. and Fierer, N. (2010) The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environmental Microbiology* **12**, 2885-2893.

Rice, W., Penney, D. and Nyborg, M. (1977) Effects of soil acidity on rhizobia numbers, nodulation and nitrogen fixation by alfalfa and red clover. *Canadian Journal of Soil Science* **57**, 197-203.

Riffkin, P.A., Quigley, P.E., Cameron, F.J., Peoples, M.B. and Thies, J.E. (1999) Annual nitrogen fixation in grazed dairy pastures in south-western Victoria. *Australian Journal of Agricultural Research*.

Roberts, A. and Morton, J. (1999) Fertiliser use on New Zealand dairy farms. *New Zealand Fertiliser Manufacturers' Association, Auckland, New Zealand Revised Edition* **37**.

Robinson, A. (1969) Competition between effective and ineffective strains of *Rhizobium trifolii* in the nodulation of *Trifolium subterraneum*. *Crop and Pasture Science* **20**, 827-841.

Rodriguez, R.J., Henson, J., Van Volkenburgh, E., Hoy, M., Wright, L., Beckwith, F., Kim, Y.-O. and Redman, R.S. (2008) Stress tolerance in plants via habitat-adapted symbiosis. *The ISME Journal* **2**, 404.

Roehr, J.T., Dieterich, C. and Reinert, K. (2017) Flexbar 3.0–SIMD and multicore parallelization. *Bioinformatics* **33**, 2941-2942.

Ronnpagel, K., Liss, W. and Ahlf, W. (1995) Microbial bioassays to assess the toxicity of solid-associated contaminants. *Ecotoxicology and Environmental Safety* **31**, 99-103.

Rousk, J., Bååth, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., Knight, R. and Fierer, N. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME Journal* **4**, 1340-1351.

Rousk, J., Brookes, P.C. and Bååth, E. (2009) Contrasting soil pH effects on fungal and bacterial growth suggest functional redundancy in carbon mineralization. *Applied and Environmental Microbiology* **75**, 1589-1596.

Rys, G. and Bonish, P. (1981) Effectiveness of *Rhizobium trifolii* populations associated with *Trifolium* species in Taranaki, New Zealand. *New Zealand Journal of Experimental Agriculture* **9**, 329-335.

Samac, D.A. and Tesfaye, M. (2003) Plant improvement for tolerance to aluminum in acid soils–a review. *Plant Cell, Tissue and Organ Culture* **75**, 189-207.

Sarita, S., Sharma, P.K., Priefer, U.B. and Prell, J. (2005) Direct amplification of rhizobial *nodC* sequences from soil total DNA and comparison to *nodC* diversity of root nodule isolates. *FEMS Microbiology Ecology* **54**, 1-11.

Scott, D. and Ronson, C. (1984) Physical Mapping of the Nif and Nod Regions of a *Rhizobium trifolii* Sym::R68. 45 Cointegrate Plasmid. In *Advances in Nitrogen Fixation Research*. pp.701-701: Springer.

Seth, K. (2017) Functional and community investigation of nodule endophytes for the selection of phosphate solubilizing rhizobial inoculants for clover: Lincoln University.

Shokralla, S., Spall, J.L., Gibson, J.F. and Hajibabaei, M. (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* **21**, 1794-1805.

Slater, F.R., Bailey, M.J., Tett, A.J. and Turner, S.L. (2008) Progress towards understanding the fate of plasmids in bacterial communities. *FEMS Microbiology Ecology* **66**, 3-13.

Slattery, J., Coventry, D. and Slattery, W. (2001) Rhizobial ecology as affected by the soil environment. *Animal Production Science* **41**, 289-298.

Smetham, M.L. (2003) A review of subterranean clover (*Trifolium subterraneum* L.): its ecology, and use as a pasture legume in Australasia. In *Advances in Agronomy*. pp.303-350: Academic Press.

Soenens, A. and Imperial, J. (2018) Novel, non-symbiotic isolates of *Neorhizobium* from a dryland agricultural soil. *PeerJ* **6**, e4776.

Sørensen, J. and Sessitsch, A. (2007) Plant-associated bacteria-lifestyle and molecular interactions. In *Modern Soil Microbiology*. pp.211-236: CRC press.

Spaink, H.P., Wijfjes, A.H., der van Drift, K.M., Haverkamp, J., Thomas-Oates, J.E. and Lugtenberg, B.J. (1994) Structural identification of metabolites produced by the NodB and NodC proteins of *Rhizobium leguminosarum*. *Molecular Microbiology* **13**, 821-831.

Sprent, J.I. (2009) *Legume Nodulation: A Global Perspective*. Chichester, UK: Wiley-Blackwell.

Stasiak, G., Mazur, A., Wielbo, J., Marczak, M., Żebracki, K., Koper, P. and Skorupska, A. (2014) Functional relationships between plasmids and their significance for metabolism and symbiotic performance of *Rhizobium leguminosarum* bv. *trifolii*. *Journal of Applied Genetics* **55**, 515-527.

Stefani, F., Isabel, N., Morency, M.-J., Lamothe, M., Nadeau, S., Lachance, D., Li, E.H., Greer, C., Yergeau, É. and Pinno, B.D. (2018) The impact of reconstructed soils following oil sands exploitation on aspen and its associated belowground microbiome. *Scientific reports* **8**, 2761.

Stowers, M.D. (1985) Carbon metabolism in Rhizobium species. *Annual Reviews in Microbiology* **39**, 89-108.

Sturz, A., Christie, B., Matheson, B. and Nowak, J. (1997) Biodiversity of endophytic bacteria which colonize red clover nodules, roots, stems and foliage and their influence on host growth. *Biology and Fertility of Soils* **25**, 13-19.

Suckling, F., Forde, M. and Williams, W. (1983) Naturalised subterranean clover in New Zealand. *New Zealand Journal of Agricultural Research* **26**, 35-43.

Sun, L.-N., Zhang, Y.-F., He, L.-Y., Chen, Z.-J., Wang, Q.-Y., Qian, M. and Sheng, X.-F. (2010) Genetic diversity and characterization of heavy metal-resistant-endophytic bacteria from two copper-tolerant plant species on copper mine wasteland. *Bioresource Technology* **101**, 501-509.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. and Willerslev, E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology* **21**, 2045-2050.

Tamura, K., Nei, M. and Kumar, S. (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences* **101**, 11030-11035.

Tan, H.W., Weir, B.S., Carter, N., Heenan, P.B., Ridgway, H.J., James, E.K., Sprent, J.I., Young, J.P.W. and Andrews, M. (2012) Rhizobia with 16S rRNA and *nifH* similar to *Mesorhizobium huakuii* but novel *recA*, *glnII*, *nodA* and *nodC* genes are symbionts of New Zealand *Carmichaelinae*. *PloS One* **7**, e47677.

Team, R.C. (2017) R: A language and environment for statistical computing.

Teixeira, C.S., Lucas, D. and Moot, D.J. (2015) Subterranean clover – Literature Review: Project Number: 408090.

Terpolilli, J., Rui, T., Yates, R., Howieson, J., Poole, P., Munk, C., Tapia, R., Han, C., Markowitz, V. and Tatiparthi, R. (2014) Genome sequence of *Rhizobium leguminosarum* bv *trifolii* strain WSM1689, the microsymbiont of the one flowered clover *Trifolium uniflorum*. *Standards in Genomic Sciences* **9**, 527.

Thies, J.E., Singleton, P.W. and Bohlool, B.B. (1991) Influence of the size of indigenous rhizobial populations on establishment and symbiotic performance of introduced rhizobia on field-grown legumes. *Applied and Environmental Microbiology* **57**, 19-28.

Unkovich, M. (2012) Nitrogen fixation in Australian dairy systems: review and prospect. *Crop and Pasture Science* **63**, 787-804.

Unkovich, M., Sanford, P. and Pate, J. (1996) Nodulation and nitrogen fixation by subterranean clover in acid soils as influenced by lime application, toxic aluminium, soil mineral N, and competition from annual ryegrass. *Soil Biology and Biochemistry* **28**, 639-648.

Unkovich, M.J. and Pate, J.S. (1998) Symbiotic effectiveness and tolerance to early season nitrate in indigenous populations of subterranean clover rhizobia from SW Australian pastures. *Soil Biology and Biochemistry* **30**, 1435-1443.

Van Overbeek, L. and Van Elsas, J.D. (2008) Effects of plant genotype and growth stage on the structure of bacterial communities associated with potato (*Solanum tuberosum* L.). *FEMS Microbiology Ecology* **64**, 283-296.

Vasileiadis, S., Puglisi, E., Arena, M., Cappa, F., Cocconcelli, P.S. and Trevisan, M. (2012) Soil bacterial diversity screening using single 16S rRNA gene V regions coupled with multi-million read generating sequencing technologies. *PloS One* **7**, e42671.

Vincent, J.M. (1970) A manual for the practical study of the root-nodule bacteria. *Blackwell Scientific Publications, Oxford*.

Voisey, C.R., White, D.W., Wigley, P.J., Chilcott, C.N., McGregor, P.G. and Woodfield, D.R. (1994) Release of transgenic white clover plants expressing *Bacillus thuringiensis* genes: An ecological perspective. *Biocontrol Science and Technology* **4**, 475-481.

Wagner, B.D., Grunwald, G.K., Zerbe, G.O., Mikulich-Gilbertson, S.K., Robertson, C.E., Zemanick, E.T. and Harris, J.K. (2018) On the Use of Diversity Measures in Longitudinal Sequencing Studies of Microbial Communities. *Frontiers in Microbiology* **9**.

Wakelin, S., Macdonald, L., Rogers, S., Gregg, A.L., Bolger, T. and Baldock, J.A. (2008) Habitat selective factors influencing the structural composition and functional capacity of microbial communities in agricultural soils. *Soil Biology and Biochemistry* **40**, 803-813.

Wakelin, S., Tillard, G., van Ham, R., Ballard, R., Farquharson, E., Gerard, E., Geurts, R., Brown, M., Ridgway, H. and O'Callaghan, M. (2018) High spatial variation in population size and symbiotic performance of *Rhizobium leguminosarum* bv. *trifolii* with white clover in New Zealand pasture soils. *PloS One* **13**, e0192607.

Watkin, E.L.J., O'Hara, G.W. and Glenn, A.R. (2003) Physiological responses to acid stress of an acid-soil tolerant and an acid-soil sensitive strain of *Rhizobium leguminosarum* biovar *trifolii*. *Soil Biology and Biochemistry* **35**, 621-624.

Watkin, E.L.J., O'Hara, G.W., Howieson, J.G. and Glenn, A.R. (2000) Identification of tolerance to soil acidity in inoculant strains of *Rhizobium leguminosarum* bv. *trifolii*. *Soil Biology and Biochemistry* **32**, 1393-1403.

Weisburg, W.G., Barns, S.M., Pelletier, D.A. and Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology* **173**, 697-703.

Wickham, H. (2011) ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics* **3**, 180-185.

Wielbo, J., Kidaj, D., Koper, P., Kubik-Komar, A. and Skorupska, A. (2012) The effect of biotic and physical factors on the competitive ability of *Rhizobium leguminosarum*. *Open Life Sciences* **7**, 13-24.

Wielbo, J., Marek-Kozaczuk, M., Kubik-Komar, A. and Skorupska, A. (2007) Increased metabolic potential of *Rhizobium* spp. is associated with bacterial competitiveness. *Canadian Journal of Microbiology* **53**, 957-967.

Wielbo, J., Marek-Kozaczuk, M., Mazur, A., Kubik-Komar, A. and Skorupska, A. (2010) Genetic and metabolic divergence within a *Rhizobium leguminosarum* bv. *trifolii* population recovered from clover nodules. *Applied and Environmental Microbiology* **76**, 4593-4600.

Wigley, K. (2017) Aspects of soil ecology of rhizobia affecting nodule occupancy on lucerne and white clover: Lincoln University.

Wigley, K., Moot, D., Wakelin, S.A., Laugraud, A., Blond, C., Seth, K. and Ridgway, H. (2017) Diverse bacterial taxa inhabit root nodules of lucerne (*Medicago sativa* L.) in New Zealand pastoral soils. *Plant and Soil* **420**, 253-262.

Williams, A., Wilkinson, A., Krehenbrink, M., Russo, D.M., Zorreguieta, A. and Downie, J.A. (2008) Glucomannan-mediated attachment of *Rhizobium leguminosarum* to pea root hairs is required for competitive nodule infection. *Journal of Bacteriology* **190**, 4706-4715.

Woodfield, D. and Easton, H. (2004) Advances in pasture plant breeding for animal productivity and health. *New Zealand Veterinary Journal* **52**, 300-310.

Xia, Y. and Sun, J. (2017) Hypothesis Testing and Statistical Analysis of Microbiome. *Genes Dis* **4**, 138-148.

Yates, R., di Mattia, E., O'Hara, G., Real, D. and Howieson, J. (2003) The role of *Rhizobium leguminosarum* biovar *trifolii* in extending (or restricting) the adaptation of annual and perennial *Trifolium* species in natural and managed ecosystems. In *New Perennial Legumes for Sustainable Agriculture* ed. Bennett, S.J. pp.116-130: University of Western Australia Press.

Yates, R., Howieson, J., Reeve, W., Brau, L., Speijers, J., Nandasena, K., Real, D., Sezmis, E. and O'hara, G. (2008) Host–strain mediated selection for an effective nitrogen-fixing symbiosis between *Trifolium* spp. and *Rhizobium leguminosarum* biovar *trifolii*. *Soil Biology and Biochemistry* **40**, 822-833.

Young, J. and Johnston, A. (1989) The evolution of specificity in the legume-*Rhizobium* symbiosis. *Trends in Ecology & Evolution* **4**, 341-349.

Young, J.P.W., Crossman, L.C., Johnston, A.W., Thomson, N.R., Ghazoui, Z.F., Hull, K.H., Wexler, M., Curson, A.R., Todd, J.D. and Poole, P.S. (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology* **7**.

Zahran, H.H. (2001) Rhizobia from wild legumes: diversity, taxonomy, ecology, nitrogen fixation and biotechnology. *Journal of Biotechnology* **91**, 143-153.

Zahran, H.H. (2017) Plasmids impact on rhizobia-legumes symbiosis in diverse environments. *Symbiosis* **73**, 75-91.

Zézé, A., Mutch, L.A. and Young, J.P.W. (2001) Direct amplification of nodD from community DNA reveals the genetic diversity of *Rhizobium leguminosarum* in soil. *Environmental Microbiology* **3**, 363-370.

Zhang, F., Lynch, D.H. and Smith, D.L. (1995) Impact of low root temperatures in soybean [*Glycine max.*(L.) Merr.] on nodulation and nitrogen fixation. *Environmental and Experimental Botany* **35**, 279-285.

Zhang, J.J., Jing, X.Y., de Lajudie, P., Ma, C., He, P.X., Singh, R.P., Chen, W.F. and Wang, E.T. (2016) Association of white clover (*Trifolium repens* L.) with rhizobia of sv. *trifolii* belonging to three genomic species in alkaline soils in North and East China. *Plant and Soil* **407**, 417-427.

Zhang, Y., Xie, J., Liu, M., Tian, Z., He, Z., van Nostrand, J.D., Ren, L., Zhou, J. and Yang, M. (2013) Microbial community functional structure in response to antibiotics in pharmaceutical wastewater treatment systems. *Water Research* **47**, 6298-6308.